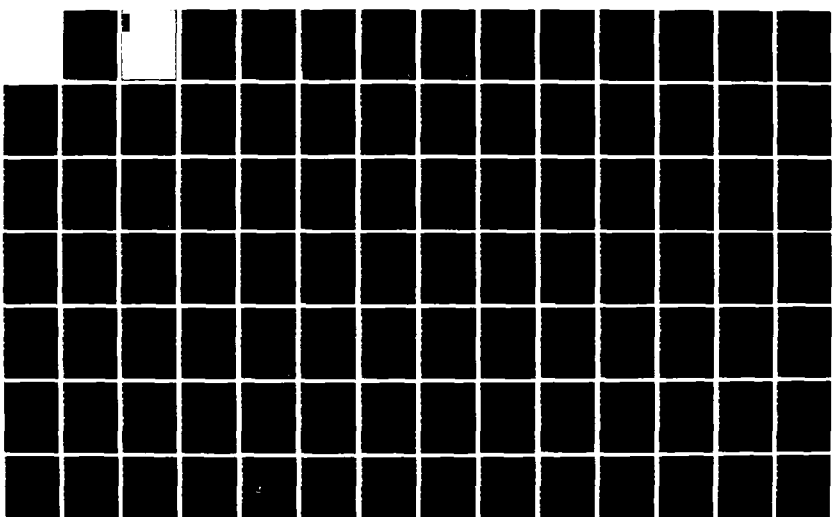
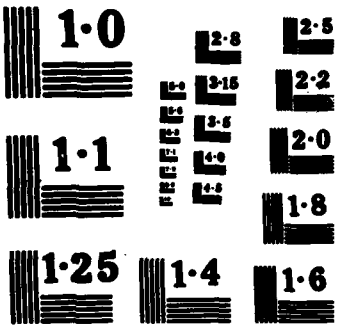
AD-A145 585 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS 1/3
AND PROGRESS OF S.. (U) HASKINS LABS INC NEW HAVEN CT
A M LIBERMAN AUG 84 SR-77/78(1984) N00014-83-K-0083
UNCLASSIFIED F/G 17/2 NL




1·0
1·1
1·25
2·8
3·15
4·0
1·4
2·5
2·2
2·0
1·8
1·6

performance across set types that infants were sensitive to the structure of consonantal segments, that is, to their particular combinations of "features."

We have then a conflict in data from the three studies: 2- to 4-month-old infants, tested with high amplitude sucking, discriminate between arbitrary sound classes that are indiscriminable for 6-month-old infants, tested with operant head-turning. If the results are valid, and not mere sampling error, we have a paradox similar to that for infants and older children with which we began. We may resolve the paradox on the same two fronts. Methodologically, we must acknowledge a commonplace of psychophysical testing for many years (e.g., Woodworth, 1938, chap. 17): different behavioral measures may give different results, even in the same individual, at roughly the same time. Moreover, since demonstrating a capacity takes precedence over demonstrating its absence, and since 6-month-old infants are unlikely to have

(11)

SR-77/78 (1984)

Status Report on

SPEECH RESEARCH

A Report on
the Status and Progress of Studies on
the Nature of Speech, Instrumentation
for its Investigation, and Practical
Applications

1 January - 30 June 1984

Haskins Laboratories
270 Crown Street
New Haven, Conn. 06511

(The information in this document is available to the general public. Haskins Laboratories distributes it primarily for library use. Copies are available from the National Technical Information Service or the ERIC Document Reproduction Service. See the Appendix for order numbers of previous Status Reports.)

Michael Studdert-Kennedy, Editor-in-Chief
Nancy O'Brien, Editor
Margo Carter, Technical Illustrator
Gail Reynolds, Word Processor

ACKNOWLEDGMENTS

The research reported here was made possible
in part by support from the following sources:

National Institute of Child Health and Human Development
Grant HD-01994
Grant HD-16591

National Institute of Child Health and Human Development
Contract NO1-HD-1-2420

National Institutes of Health
Biomedical Research Support Grant RR-05596

National Science Foundation
Grant BNS-8111470

National Institute of Neurological and Communicative
Disorders and Stroke
Grant NS 13870
Grant NS 13617
Grant NS 18010

Office of Naval Research
Contract N00014-83-K-0083

HASKINS LABORATORIES PERSONNEL IN SPEECH RESEARCH

Alvin M. Liberman,* President
Franklin S. Cooper,* Assistant to the President
Patrick W. Nye, Vice President
Michael Studdert-Kennedy,* Vice President, Research
Raymond C. Huey,* Treasurer
Bruce Martin, Controller
Alice Dadourian, Secretary

Investigators

Arthur S. Abramson*
Peter J. Alfonso*
Thomas Baer
Patrice S. Beddor†
Fredericka Bell-Berti*
Shlomo Bentin[1]
Catherine Best*
Gloria J. Borden*
Susan Brady*
Catherine P. Browman
Robert Crowder*
Laurie B. Feldman*
Carol A. Fowler*
Louis Goldstein*
Vicki L. Hanson
Katherine S. Harris*
Sarah Hawkins††
Satoshi Horiguchi[2]
Leonard Katz*
J. A. Scott Kelso
Andrea G. Levitt[3]
Isabelle Y. Liberman*
Leigh Lisker*
Anders Löfqvist[4]
Virginia Mann*
Ignatius G. Mattingly*
Nancy S. McGarr*
Robert J. Porter, Jr.[5]
Lawrence J. Raphael*
Bruno H. Repp
Philip E. Rubin
Elliot Saltzman
Donald Shankweiler*
Mary Smith*
Betty Tuller*
Michael T. Turvey*
Ben C. Watson††
Douglas H. Whalen

Technical/Support

Michael Anstett
Margo Carter
Philip Chagnon
Vincent Gulisano
Donald Hailey
Sabina D. Koroluk
Betty J. Myers
Nancy O'Brien
Gail K. Reynolds
William P. Scully
Richard S. Sharkany
Edward R. Wiley

Students*

Eric Bateson
Suzanne Boyce
Jo Ann Carlisle
Andre Cooper
Patricia Ditunno
Jan Edwards
Jo Estill
Nancy Fishbein
Carole E. Gelfer
Charles Hoequist
Bruce Kay
Noriko Kobayashi
Rena A. Krakow
Harriet Magen
Sharon Manuel
Richard McGowan
Daniel Recasens
Lawrence D. Rosenblum
Hyla Rubin
Judith Rubin-Spitz
Richard C. Schmidt
John Scholz
Robin Seider
Suzanne Smith
Katyanee Svastikula
Daniel Weiss
Deborah Wilkenfeld
David Williams

*Part-time
[1]Visiting from Hadassah University Hospital, Jerusalem, Israel
[2]Visiting from University of Tokyo, Japan
[3]On leave, Centre d'Etude Processus Cognitifs et du Langage, Paris, France
[4]Visiting from Lund University, Lund, Sweden
[5]Visiting from University of New Orleans and Kresge Research Laboratory of the South, New Orleans, Louisiana
†NIH Research Fellow
††NRSA Training Fellow

CONTENTS

SOURCES OF VARIABILITY IN EARLY SPEECH DEVELOPMENT*

Michael Studdert-Kennedy†

The present paper considers the origins of differences among children, and within a child from time to time, in the early development of speech. The bias of the paper is toward viewing these differences as special cases of general variability in animal behavior and its development. Some variability among children is surely genetic in origin (Lieberman, this volume); this is the stuff of natural selection. Other variability is precisely what we expect in a system growing from an open genetic program (Mayr, 1974) that depends on loosely invariant properties of the environment to specify the course of development (for elaboration, see below, and for an excellent brief discussion, see Lenneberg, 1967, chap. 1). Finally, variability within a child is a precondition of the adaptive biological process that we term "learning" (cf. Fowler & Turvey, 1978). However, I will come to all these matters only in the last section of the paper.

My first concern, and the topic of the early parts of the paper, is apparent differences between capacities of infants and older children. Ferguson (this volume) notes two main areas of research in child phonology: speech perception in infants, and the sound systems of individual children aged 2-4 years, as shown by their speech productions. The relation between these two bodies of work is, indeed, "problematic," as Ferguson remarks. For, on the one hand, we have an infant apparently capable not only of discriminating virtually every adult segmental contrast with which it is presented, but also of discriminating speech sound categories across speakers and perhaps even across intrinsic allophonic variants (for a comprehensive review, see Aslin, Pisoni, & Jusczyk, 1983). On the other hand, we have an older child producing a bewildering variety of sounds in its attempts to reproduce a particular adult word. The discrepancy is not simply between perception and production. For we also find the older child, even up to the age of 5 or 6 years, making substantial numbers of perceptual errors on consonant contrasts (voicing, nasality, place of articulation) that would, seemingly, have caused no difficulty at all when it was an infant (see Barton, 1980, for a review). Of course, these are cross-sectional comparisons. But the data are well established, and would usually be taken to reflect the child's course of development rather than sampling error.

---

*To appear in J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: Erlbaum, in press.

†Also Queens College and Graduate Center, City University of New York.

Acknowledgment. My thanks to Björn Lindblom and Peter MacNeilage for conversations, to Charles Ferguson and Lise Menn for their papers, to John Locke for his book, and my apologies to all of them for any misconstruals. Preparation of the paper was supported in part by NICHD Grant No. HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

How then are we to resolve the paradox? The first step is to acknowledge that different tasks place different demands on infant and older child: to detect the difference between two patterns of sound (discrimination) is not necessarily to recognize each pattern as an instance of a category (identification) (Barton, 1980, p. 106). Moreover, even when the tasks assigned to infant and older child are the same (i.e., discrimination), different behavioral measures may give different results: recovery from habituation to a nonsense syllable upon presentation of a new syllable, as measured by high amplitude sucking or by heart rate, may not draw on the same capacities as choosing which of two nonsense words refers to a particular wooden block (Garnica, 1971). If we assume, as seems reasonable, that the older child has not lost capacities for discriminating between sounds of the surrounding language that it possessed as an infant, we must conclude that those capacities are not sufficient for more explicitly communicative tasks (cf. Oller & Eilers, 1983; Oller & MacNeilage, 1983).

Yet the origin of the paradox is more than methodological. It also arises because infant speech research has "...generally taken for granted a phonological unit corresponding to the 'segment' [or, we may add, feature] of contemporary phonological theories, even though researchers have sometimes been familiar with the problems of relating such abstract units to the processes of speech perception..." (Ferguson, this volume). Ferguson himself has a different and, I believe, more fruitful approach. For rather than viewing the child as "acquiring" its phonology from the adult, Ferguson sees the adult's phonology as growing out of the child's (cf. Locke, 1983; Menyuk & Menn, 1979). Moreover, like Moskowitz (1973), and in accord with sound biological principle (e.g., Waddington, 1966), Ferguson sees this growth as a process of differentiating smaller structures from larger. The child does not build words with phonemes: phonemes emerge from words. In short, Ferguson shuns the preformationist view (long banished from embryology, but still thriving in psychology) that attributes adult properties to the child; he seeks rather to trace the epigenetic course from child to adult.

In the next few sections I will sketch a view of infant speech development over roughly the first year of life that attempts to resolve the "problematic" relation between the apparent capacities of infant and older child. Broadly, my view is that two wrong turns have led into the impasse. First, a too narrow notion of development has encouraged undue concentration on the infant's "initial state." For the biologist, development begins with the first division of the fertilized egg and ends with death. At each moment, the organism is sufficient for adaptive response to current internal and external conditions. Birth is certainly an occasion of abrupt discontinuity and of radical changes in conditions, but prenatal and postnatal development do not differ in principle: the infant's state at birth is simply the first state that psychologists can conveniently study.

Of course, we may treat the whole process teleologically, seeing the end in the beginning. That, in my view, is the second wrong turn. For the habit of describing infants' presumed percepts (and articulations) in linguistic terms has diverted attention from the central problem of early speech development, namely, imitation. We have been easily diverted because it seems natural (as, indeed, it is) that, if an adult speaks a word or grasps the air with her hand, a young child can repeat the word or imitate the hand movements. But how, in fact, does the child do this? What information in the acoustic or optic array specifies the executed movements? How is the information trans-

duced into muscular controls? We are far from even imagining an answer to the last question. But we may gain leverage on the former (the very question to which the infant, learning to speak, must itself find an answer), if we couch our descriptions in auditory and motoric, rather than in linguistic, terms. We begin then with a brief summary of what is known about speech perceptuomotor processes in adults.

Cerebral Asymmetry for Language in Adults

Brain lateralization offers a chink through which we may view the early stages of imitative processes essential to language development. To justify this claim my first assumption is that the association between lateralizations for language and manual praxis in more than 90% of the human population (Levy, 1974) is not mere coincidence. Second, I assume that lateralization of hand control evolved in higher primates to facilitate bimanual coordination by assigning unilateral control to a bilaterally innervated system (MacNeilage, Studdert-Kennedy, & Lindblom, 1984). Third, I assume that speech and language exploited the already existing neural organization of the left hemisphere to develop a characteristic structure, analogous in certain key respects to the structure of coordinated hand movements.

I have no space to develop the analogy here (for elaboration, see MacNeilage, 1983; MacNeilage, Studdert-Kennedy, & Lindblom, in press). In any case, for present purposes, the needed assumption is simply that language evolved in the left hemisphere for reasons of motor control. The assumption is consistent with studies of aphasics (Milner, 1974), of split-brain patients (Zaidel, 1978) and of the effects of sodium amytal injection (Borchgrevink, 1983; Milner, Branch, & Rasmussen, 1964), showing that in most right-handed individuals the right hemisphere is essentially mute: the bilaterally innervated speech apparatus is controlled from the left side.

My final assumption is that a capacity to perceive speech--more exactly, to break its patterns into components matched to the motor components of articulation--evolved alongside the motor system in the left hemisphere. The assumption is consistent with numerous studies of dichotic listening (e.g., Kimura, 1961, 1967; Studdert-Kennedy & Shankweiler, 1970), and has drawn further support from studies of split-brain patients. Levy (1974) showed that only the left hemisphere of these patients can carry out the phonological analysis needed to recognize written rhymes; Zaidel (1976, 1978) showed that, while the right hemisphere may have a sizeable auditory and visual lexicon, only the left hemisphere can carry out the auditory-phonetic analysis necessary to identify synthetic nonsense syllables, or the phonological analysis necessary to read new words.

In short, the stated assumptions and their supporting evidence justify the claim that the speech perceptuomotor system is vested in the left hemisphere of most normal right-handed individuals. Let us turn now to the development of this system over the first year of life.

Cerebral Asymmetry for Speech in Infants

Perception. A number of perception studies has demonstrated dissociation of the left and right sides of the brain for perceiving speech and non-speech sounds at, or very shortly after, birth. For example, Molfese, Freeman and Palermo (1975) measured auditory evoked responses, over left and right tempo-

ral lobes, of 10 infants, ranging in age from one week to 10 months. Their stimuli were four naturally spoken monosyllables, a C-major piano chord and a 250-4000 Hz burst of noise. Each stimulus lasted 500 ms and was presented about 100 times, at randomly varying intervals. Median amplitude of response was higher over the left hemisphere for all four syllables in nine out of ten infants, higher over the right hemisphere for the chord and the noise in all ten infants; the one child who responded to speech with higher right hemisphere amplitude had a left-handed mother. Molfese (1977) has reported similar asymmetries for syllables and pure tones in neonates.

Segalowitz and Chapman (1980) studied 153 premature infants with a mean gestational age at testing of 36 weeks. They measured reduction of limb tremor over a 24-hour period, at the end of a daily regimen of exposure to 5-minute spells of speech (the mother reading nursery rhymes) or music (Brahms' "Lullaby"), presented six times a day at 2-hour intervals. Tremor in the right arm (but not in the right leg, nor in the left arm or leg) was significantly more reduced by speech than by music or by silence (control group). The mechanism of the effect is not understood, nor whether it is due to cortical or subcortical asymmetries.

Finally, Best, Hoffman, and Glanville (1982) tested forty-eight 2-, 3- and 4-month old infants for ear differences in a memory-based dichotic task. They used a cardiac orienting response to measure recovery from habituation to synthetic stop-vowel syllables and to Minimoog simulations of concert A (440 Hz) played on different instruments. In the speech task, a single dichotic habituation pair (either /ba-da/ or /pa-ta/) was presented nine times, at randomly varying intervals. On the 10th presentation, one ear again received its habituation syllable, while the other received a test syllable (either /ga/ or /ka/), differing in place of articulation from both habituation syllables. An analogous procedure was followed in the musical note task.

The results showed significantly greater recovery of cardiac response for right ear test syllables in the 3- and 4-month-olds, and for left ear musical notes in all age groups. The authors suggest that right-hemisphere memory for musical sounds develops before left-hemisphere memory for speech sounds, and that the latter begins to develop between the second and third months of life.

Neither these nor any of the several other studies with similar findings (see Best et al., 1982, for a brief review) indicate what properties of the signal mark it as speech. We may note, however, that those properties are evidently present in isolated syllables, natural or synthetic, and do not depend on the melody or rhythm of fluent speech. Moreover, the results of Best et al. (1982) invite the inference that infant speech sound discrimination, attested by numerous studies, engages left-hemisphere mechanisms no less than does adult speech sound discrimination.

Production. Evidence for early development of the production side of the perceptuomotor link is tenuous, but suggestive. Kuhl and Meltzoff (1982) showed that 4- to 5-month-old infants looked longer at the video-displayed face of a woman articulating the vowel they were hearing (either [i] or [a]) than at the same face articulating the other vowel in synchrony. The preference disappeared when the signals were pure tones matched in amplitude and duration to the vowels, so that infant preference was evidently for a match between mouth shape and spectral structure. Similarly, MacKain, Studdert-Kennedy, Spieker, and Stern (1983) showed that 5- to 6-month-old infants pre-

ferred to look at the face of a woman repeating the disyllable they were hearing (e.g., [zuzi]) than at the synchronized face of the same woman repeating another disyllable (e.g., [vava]). In both these studies infant preferences were for natural structural correspondences between acoustic and optic information. Since these two sources of information have a common origin in the articulations of the speaker, we may reasonably infer that the infant is sensitive to information that specifies articulation. (For related work on adult "lip-reading," see Campbell & Dodd, 1979; Crowder, 1983; McGurk & MacDonald, 1976; Summerfield, 1979).

Two more items complete the circle. First, Meltzoff and Moore (1977) showed that 12- to 21-day-old infants could imitate both arbitrary mouth movements, such as tongue protrusion and mouth opening, and (of interest for the development of ASL) arbitrary hand movements, such as opening and closing the hand by serially moving the fingers. Here mouth opening was elicited without vocalization; but had vocalization occurred, its structure would necessarily have reflected the shape of the mouth. Kuhl and Meltzoff (1982) do, in fact, report as an incidental finding of their study that 10 of their 32 infants "...produced sounds that resembled the adult female's vowels. They seemed to be imitating the female talker, 'taking turns' by alternating their vocalizations with hers" (p. 1140). Of course, we have no indication that this incipient capacity, demonstrated under conditions of controlled attention in the laboratory, is actively used by 5-month-old infants in the more variable conditions of daily life.

The second item of evidence is a curious aspect of the study by MacKain et al. (1983), cited earlier: infant preferences for a match between the facial movements they were watching and the speech sounds they were hearing were statistically significant only when they were looking to their right sides. Fourteen of the eighteen infants in the study preferred more matches on their right sides than on their left. Moreover, in a follow-up investigation of familial handedness, MacKain and her colleagues learned that six of the infants had left-handed first- or second-order relatives. Of these six, four were the infants who displayed more left-side than right-side matches.

These results can be interpreted in the light of work by Kinsbourne and his colleagues (e.g., Kinsbourne, 1972; Lempert & Kinsbourne, 1982). This work suggests that attention to one side of the body may facilitate processes for which the contralateral hemisphere is specialized. If this is so, we may infer that infants with a preference for matches on their right side were revealing a left hemisphere sensitivity to articulations specified by acoustic and optic information. Thus, we have preliminary evidence that 5- to 6-month-old infants, close to the onset of babbling, already display the beginnings of a speech perceptuomotor link in the left hemisphere.

Here we should strike a note of caution. The evidence reviewed up to this point does not demonstrate that specialized phonetic processes are occurring in the infant. In fact, whatever mechanisms for imitating articulation may be developing in these early months seem to be no different, in principle, than corresponding specialized mechanisms for imitating movements of hand, face, and body. What distinguishes the speech perceptuomotor link, at this stage of development, is, first, its locus in the brain, and second, its modality. The capacity to imitate vocalizations seems to be peculiar to certain birds, certain marine mammals, and man.

Speech Perception in Infants

0-6 months.[1] As we have already remarked, and as is well known, infants in the first six months of life discriminate almost any adult segmental contrast on which they are tested. Particularly striking, in the early years of this work, initiated by Eimas and his colleagues (Eimas, Siqueland, Jusczyk, & Vigorito, 1971), was 1- and 4-month-old infants' discrimination of synthetic syllables along a stop consonant voice-onset time continuum. Discrimination was measured by recovery (or no recovery) of high-amplitude sucking on a non-nutritive nipple, in response to a change in sound (or no change for a control group), after habituation to repeated presentation of another sound. Like adults, infants readily discriminated between acoustically different items belonging to different (English) phonetic categories, but not between acoustically different items belonging to the same category. This finding, fortified by similar results on continua of, for example, stop consonant place of articulation (Eimas, 1974), consonant manner (Eimas & Miller, 1980a, 1980b), and the [r]-[l] distinction (Eimas, 1975), encouraged the hypothesis that "...these early categories serve as the basis for future phonetic categories" (Eimas, 1982, p. 342).

However, there is a confusion here between two different types of category. On the one hand, we have categories comprising more-or-less random variations in the precise acoustic properties of a single syllable, spoken repeatedly with identical stress and at an identical rate by the same speaker: these are the patterns mimicked by a synthetic series, varied along a single acoustic dimension. On the other hand, we have the categories of natural speech, comprising intrinsic allophonic variants, formed by the execution of a particular phoneme in a range of phonetic contexts, spoken with varying stress, at different rates and by different speakers. The latter are presumably the "future phonetic categories" to which Eimas refers, while the former are auditory categories to which infants, chinchillas (for VOT: Kuhl & Miller, 1978) and macaques (for place of articulation: Kuhl & Padden, 1983) have been shown to be sensitive in synthetic speech studies (see also Kuhl, 1981). The proper interpretation of these studies would seem then to be that infants (and an open set of other animals) can discriminate the several contrasts tested, if they are presented in an invariant acoustic context.

Evidence for "phonetic" categories from studies of contrasts across varying acoustic contexts differs depending on the nature of the variation. Talker variations, at least on the few contrasts that have been tested, seem to cause little difficulty for infant (e.g., Hillenbrand, 1983; Kuhl, 1979) dog (Baru, 1975), cat (Dewson, 1964) or chinchilla (Burdick & Miller, 1975). Cross-talker categories, then, seem to be auditory rather than phonetic. (We may note, in passing, that such findings present a puzzle for accounts of speaker normalization that rest on the listener's presumed knowledge of the speaker's phonetic space [e.g., Gerstman, 1968; Ladefoged & Broadbent, 1957].)

Studies of contrasts across variations in phonetic context have given less consistent results. Warfield, Ruben, and Glackin (1966) trained cats to discriminate between the words cat and bat, but found no transfer of training to other minimal pairs, beginning with the same segments. Holmberg, Morgan and Kuhl (1977) studied fricative perception in 6-month-old infants. They used an operant head-turning paradigm, in which the infant was conditioned to turn its head for visual reinforcement when repeating sounds from one category

were changed to repeating sounds from another. They found that infants discriminated [f]/[θ] and [s]/[ʃ] across variations in vowel context (e.g., [fa], [fi], [fu]) and syllable position (e.g., [fa], [af]). Kuhl (1980) reports similar results for an infant, trained to discriminate [d]/[g].

Katz and Jusczyk (1980), cited in Jusczyk (1982), reasoned that a more stringent test of infant phonetic categorization would be to show that infants more readily learn to discriminate between (that is, to generalize within) phonetically-based groupings than arbitrary groupings of the same syllables. In a head-turning study of 6-month-old infants, they found that most infants learned to discriminate between sets of syllables, paired for consonant onset, but differing in vowel (e.g., [bi] and [bɛ] vs. [di] and [dɛ]), but not between sets, arbitrarily paired, differing in both consonant and vowel (e.g., [bɛ] and [di] vs. [bi] and [dɛ]). However, none of the infants learned to discriminate either phonetic or arbitrary groupings of [b] and [d] followed by four vowels ([i, ɛ, o, ɚ]). Jusczyk (1982) interprets the results as providing some "...weak support for...perceptual constancy for stop consonant segments occurring in different contexts" (p. 378).

Before commenting on this study, let us compare its results with those of Miller and Eimas (1979), who used a similar set of stimulus materials, to ask a different experimental question: Are infants sensitive to the structure of syllables? That is to say, do infants perceive syllables holistically, as seamless, undifferentiated patterns, or do they perceive the structure of syllables, analyzing them into their component segments (consonants and vowels)? Miller and Eimas used a high-amplitude sucking paradigm to test 2-, 3-, and 4-month-old infants. One group of infants successfully discriminated between sets of syllables, paired for consonant onsets, but differing in vowel ([ba] and [bæ] vs. [da] and [dæ]), as did the infants of Katz and Jusczyk. However, another group also discriminated between sets arbitrarily paired, differing in both consonant and vowel ([ba] and [dæ] vs. [bæ] and [da]), as the infants of Katz and Jusczyk did not. Miller and Eimas interpreted their positive outcome as evidence that infants are sensitive to the segmental structure of syllables.

A similar conflict in results emerges at a "feature" level when we compare a study by Hillenbrand (1983) with the second and third experiments of Miller and Eimas (1979). Hillenbrand used a head-turning paradigm to test the capacity of 6-month-old infants to discriminate between sets of syllables differing on a single feature (oral-nasal, as in [ba] and [da] vs. [ma] and [na]) and sets of syllables differing on arbitrary combinations of two features (oral-nasal and place of articulation, as in [ba] and [ŋa] vs. [na] and [ga]). He found that infants were significantly better at discriminating the single feature "phonetic" groups than the arbitrary double feature groups. He concluded that infants were sensitive to the auditory correlates of consonantal features. Miller and Eimas (1979), on the other hand, in two further experiments of their study, tested 2- , 3- and 4-month-old infants, with a high amplitude sucking procedure, on single-feature phonetic groups analogous to those of Hillenbrand (voicing vs. place of articulation, oral-nasal vs. place of articulation), and on the corresponding double feature sets where the two "features" were arbitrarily combined. Pooling data from the two experiments, they found that infants assigned to experimental conditions displayed significantly more recovery from habituation than control infants, and that there was no significant difference in recovery for the two types of syllable set. Miller and Eimas (1979) concluded from the lack of reduction in

performance across set types that infants were sensitive to the structure of consonantal segments, that is, to their particular combinations of "features."

We have then a conflict in data from the three studies: 2- to 4-month-old infants, tested with high amplitude sucking, discriminate between arbitrary sound classes that are indiscriminable for 6-month-old infants, tested with operant head-turning. If the results are valid, and not mere sampling error, we have a paradox similar to that for infants and older children with which we began. We may resolve the paradox on the same two fronts. Methodologically, we must acknowledge a commonplace of psychophysical testing for many years (e.g., Woodworth, 1938, chap. 17): different behavioral measures may give different results, even in the same individual, at roughly the same time. Moreover, since demonstrating a capacity takes precedence over demonstrating its absence, and since 6-month-old infants are unlikely to have lost capacities for discriminating among the sounds of the surrounding language that they possessed at 3 months, we must conclude that high-amplitude sucking is a more sensitive measure of infant discriminative capacity than operant head-turning. Thus, the two head-turning studies failed to reveal infant conditioning to arbitrary groupings of syllables because task difficulty and behavioral measure interacted--a possibility raised by Jusczyk (1982, p. 379).[2] The attempt to develop a more stringent test of infant consonant categorization across vowel contexts than that used by Holmberg et al. (1977) for fricatives was therefore not successful.

Beyond the methodological issue lies the matter of interpretation. Consider, first, the conclusion from Miller and Eimas (1979) that infants are sensitive to the segmental structure of syllables and the featural structure of segments. Unfortunately, the conclusion is not forced by the data, since, as Aslin et al. (1983) point out, an infant discriminating, say [ba] and [na] from [da] and [ma], has simply to detect that one (or both) of the syllables in the second set is different from either of the syllables in the first set. In other words, the infant can discriminate the patterns holistically without analysis. Miller and Eimas (1979) recognize this fact ("...we know of no way to make this distinction [holistic/analytic] experimentally with infant subjects"), but justify their preference for the analytic interpretation, because "There is...rather extensive behavioral as well as neurophysiological evidence for an analysis into components or features in human and non-human pattern perception" (both quotations from p. 355, footnote 2). I do not doubt this evidence, but it does not justify our attributing analytic capacities to the 3-month-old--particularly when, by doing so, we set up a paradoxical discrepancy between the capacities of infant and older child.

Consider, next, the evidence that infants can form "phonetic" categories across a variety of acoustic contexts. Here again the data are overinterpreted. For, in fact, since every phonetic contrast is marked by an acoustic contrast (if it were not, how would the infant learn to talk?), phonetic and auditory perception cannot be dissociated in the infant (though they can be in the adult: Best, Morrongiello, & Robson, 1981; Best & Studdert-Kennedy, 1983; Liberman, Isenberg, & Rakerd, 1981; Mann & Liberman, 1983; Schwab, 1981). This fact is recognized by Miller and Eimas (1979, p. 365), and by Aslin et al. (1983, passim). What we are left with then is evidence that infants, in their first six months of life, can detect auditory similarities across certain adult phonetic categories. Incidentally, apart from the study of cats mentioned above (Warfield et al., 1966), we have no evidence, so far as I know, that other animals cannot do the same. Of course, proving the null hypothesis on animals is a thankless task.

Finally, we may ask what role categories, whether auditory or phonetic, are presumed to play in the infant's learning to speak. Eimas (1982) argues that "...the acquisition of the complex rule systems of linguistics requires that the young child treat all instantiations of a phonetic category as members of a single equivalence class" (p. 346). He adds in a footnote: "...if the child treats each possible member of the two voicing categories of English as separate entities and not as perceptually identical events or at least as members of the same equivalence class, then acquisition of the rule for pluralization will necessarily be painfully slow, if ever learned" (p. 346, footnote 5). Eimas goes on to justify the search for perceptual constancy in infants on grounds of parsimony, because "...it would effectively eliminate explanations based on receptive experience" (p. 346).

There are several things wrong here. First is the implication that a process of development relying on experience to direct its course is somehow unparsimonious, perhaps even not "biological." In fact, just the reverse is true. Precisely because full genetic specification is costly, even the lowliest behaviors of non-human animals may depend on broadly invariant external conditions to guide development. (see Immelmann, Barlow, Petrinovich, & Main, 1981, passim; Lenneberg, 1967, chap. 1; Mayr, 1974, and the brief discussion below). Second, the notion of rule is prescriptive, as though speakers applied rules much as they do in a game of chess. In fact, a phonological rule is simply a description of regularities in speech; the processes by which these regularities arise are completely unknown (for excellent discussions, see Menn, 1980; Menyuk & Menn, 1979). Finally, once again, the outcome of development (the formation of phonological structures that control adult speaking) is posited to be already in place at a time when development has scarcely begun. I do not doubt that infants can form auditory categories, but there is no evidence that this capacity is either needed for or brought to bear on early speaking.[3] If it were, we would be hard put to explain the word-by-word development of adult phones that Ferguson (this volume) describes, or the relatively slow accumulation of the first fifty (or so) words. We may indeed suspect that the emergence of auditory-motoric categories, around the beginning of the third year, is a factor in triggering the explosive growth of the child's vocabulary (at an average rate of perhaps 5-10 words a day) over the next four or five years (Miller, 1977, pp. 150 ff.).

In short, we can resolve the paradoxical discrepancy between the capacities of infants and older children, if we refrain from regarding precursors of a behavior as instances of the behavior itself. No doubt, infant kicking and stepping (when held erect) are precursors of walking and, with normal growth in an appropriate environment, will develop into walking (Thelen, 1983). But infant kicks and steps are not strides.

7-12 months. None of the foregoing should be interpreted as claiming that phonetically relevant development of the infant's perceptual system is not going forward during the first six months of life. However, the first (and still sparse) behavioral evidence of such development comes from older infants.

Eimas (1975) showed that 4- to 6-month-old English infants discriminated between English [r] and [l]. On the assumption that Japanese infants would have done the same, and given the well-known fact that native Japanese speakers, who know no English, do not make this discrimination (Miyawaki et al., 1975), Eimas suggested that learning the sound system of a language may entail

loss of the capacity to discriminate contrasts not used in the language. Similar suggestions have been made by Aslin and Pisoni (1980), Locke (1983), and others.

Werker and her colleagues (1981) have traced the onset of perceptual loss to the second six months of life, a period when the infant is perhaps first attending to individual words and the situations in which they occur (cf. Jusczyk, 1982; MacKain, 1982). Their initial finding was that seven-month-old Canadian English infants, tested in a head-turning paradigm, could discriminate between naturally spoken contrasts in Hindi as English-speaking adults could not. Werker (1982) followed this up by tracking the decline of discriminative capacity in cross-sectional and longitudinal studies. She used a conditioned head-turning paradigm to test three groups of infants on two non-English sound contrasts: Hindi voiceless, unaspirated retroflex vs. dental stops (cf. Locke, 1983, pp. 90-92), and Thompson (Interior Salish, an American Indian language) voiced, glottalized velar vs. uvular stops. On the Hindi contrast, the number of infants successfully discriminating were: 11/12 at 6-8 months, 8/12 at 8-10 months, 2/10 at 10-12 months; for the Thompson contrast the results were essentially the same. (An infant was classified as having failed to discriminate only if it had successfully discriminated an English contrast both before and after failure on a non-English contrast). Finally, Werker (1982) reports longitudinal data for six Canadian English infants on the same two non-English contrasts. All six discriminated both contrasts at 6-8 months, but at 10-12 months none of them made the discrimination. By contrast, the one Thompson and two Hindi infants so far tested at 10-12 months could all make the called for discrimination in their own language.

Perceptual loss is not permanent, since capacity can be recovered by adults learning a new language (e.g., MacKain, Best, & Strange, 1981). Nor can the effect be general, since sufficiently salient foreign contrasts can presumably be discriminated even by adults. We may suspect then that loss is focused on relatively fine auditory contrasts, specifying slight differences in the space-time coordinates of a single articulator's movements, and that it arises as a side-effect (lateral inhibition!) of the infant's developing "attention" to closely related contrasts in its own language. This is not to suggest that the younger infant is not "attending" to speech during its early months. Rather, its search for meaning and communicative function (Trevarthen, 1979) may initially be guided by the rhythm and melody of speech (e.g., Mehler, Barrière, & Jasik-Gerschenfeld, 1976). Only when these larger patterns have begun to take form (Menn, 1978a), are the infant's capacities for segmental discrimination, readily demonstrated in the laboratory, brought to bear on the speech it hears at home.

Speech Production in the Infant

The infant, by definition, does not speak (Latin: infans, not speaking). But there is now ample evidence that the discontinuity between babble and speech, posited by Jakobson (1968), is not real. Oller (1980) provides a taxonomy of the emerging stages from phonation (0-1 month) to variegated babbling (11-12 months). Oller, Wieman, Doyle, and Ross (1975) describe similarities between patterns of babbling and early speech (cf. MacNeilage, Hutchinson, & Lasater, 1981). Vihman, Macken, Miller, Simmons, and Miller (in press) demonstrate parallels in the distribution and organization of sounds in speech and babble during the period (roughly 9-15 months) when they overlap.

What is the origin of this continuity? The first possibility is that the sound distributions of babble and early speech are similar because the infant begins to learn the sounds of the language around it and to practice them during its second six months of life. Locke (1983, chap. 1) has marshalled evidence against this view. First, he has collated data on the babbling of 9- to 12-month-old infants growing up in 14 different language environments, distributed across some half dozen language families (Locke, 1983, Table 1.3, p. 10). These infants were certainly old enough to have begun to discover the sound patterns of their languages and, indeed, if the data on perceptual loss reviewed above have any generality, perceptual discovery had already begun. Yet of the 143 consonantal sounds entered in Locke's table over 85% correspond to one of the twelve most frequent sounds in the babbling of English children: a strikingly homogeneous distribution. Second, Locke has reviewed some dozen studies that have looked for drift in the sounds of infant babbling, during the second six months of life, toward the sounds of the surrounding language. Most of the studies either found no evidence of drift or were inconclusive. Finally, Locke has reviewed available studies on the babbling of deaf and Down's syndrome infants. Despite the common belief that deaf babbling fades before the end of the first year, several studies agree that it may continue well into early childhood (5-6 years). But what is remarkable is that the developmental course of babbling up to 12 months is similar in deaf and hearing infants, and, incidentally, in Down's syndrome infants. For example, the relative proportions of labial, alveolar and velar consonants follow essentially the same course: only after the 12th month does the expected preponderance of labial movements in deaf children begin. The three strands of evidence converge on a process of articulatory development, independent of the surrounding language and common to all human infants.

We are left, then, with the second possible account of the continuity between babble and speech, namely that, as Locke proposes, the phonetic proclivities of adults and infants are similar. Both are largely determined by anatomical and physiological constraints on the signaling apparatus. What these constraints may be has only recently come under scrutiny (e.g., Kent, 1980; Lindblom, 1983; Ohala, 1983).

Of course, this hypothesis raises immediately the question of language change: if all adult speakers develop from a common infant base, why do languages differ? The question is too large, and my competence too small, for adequate treatment here. However, I note several points. First, as Locke (1983) has shown, many infant biases (e.g., for open rather than closed syllables, for stops over fricatives, for singleton consonants over clusters, and so on) are indeed preserved by many groups of adult speakers (i.e., languages), and it is this fact that the continuity of babble and speech reflects. At the same time, infant preferences are not rigid, because, as Darwin taught, no animal structure specifies a unique function: A structure (e.g., the vocal apparatus) permits an unspecifiable, though presumably limited, range of functions, and the natural variability of behavior offers this range for selection. Second, infant articulatory capacities are a subset of the capacities of mature speakers. As skill develops, the range of response, available for selection by a variety of sociocultural forces, widens. Certainly, the exact course of historical change will never be fully specified for language, any more than for, say, clothing, cuisine, or social organization. Nonetheless, there would seem to be no reason, in principle, why we should not develop a cultural-evolutionary account of language diversity (Lindblom, 1984), compatible with relatively fixed infant articulatory proclivities.

The conclusion I want to draw, then, is that perceptual and motor development of speech over the first year of life, as manifested in infant behavior, may justly be seen as parallel, independent processes. No doubt, physiological changes in the perceptual and motor centers of the left hemisphere are taking place to prepare for the ultimate linkage between the two systems. These processes may be analogous to those in songbirds, such as the marsh wren, in which the perceptual template of its species' song is laid down many months before it begins to sing (Kroodsma, 1981). But behavioral evidence of the perceptuomotor link appears only with that song, just as behavioral evidence of the link appears in the infant only with its first imitation of an adult sound.

From Babbling to Speech

The transition from babbling to speech is a murky period. At this stage we see the first clear evidence of a perceptuomotor link, but know little about what the child perceives. Even when the perceptual data come in, it will be a delicate task to determine their relevance. For as we have noted, a capacity demonstrated in the laboratory does not tell us how, or even if, that capacity is put to use in learning to speak. Consequently, we may have to place as much weight on shaky inference from the child's productions as on firm evidence from perceptual studies.

A further difficulty at this stage is that we find it increasingly difficult to refrain from describing the child's productions by means of phonetic transcriptions. Of course, we do not want to refrain: transcription is our readiest mode of description, because children have vocal tracts very like adults' and make sounds like adults' sounds. Yet transcription is a double-edged blade. For it is precisely in order to understand the apparently segmented structure of speech (and the resulting adult capacity to transcribe) that we are studying its ontogeny. As is well known, phonetic segments are not readily specified either in articulation or in the signal, so that their functional reality has had to be inferred, in the first instance, from adult behaviors, such as errors of perception (e.g., Browman, 1978) and production (e.g., Shattuck-Hufnagel, 1983), backward talking (Cowan, Leavitt, Massaro, & Kent, 1982), aphasic deficits (e.g., Blumstein, 1981) and, not least, use of the alphabet. By relying on a descriptive apparatus that derives from characteristics of mature speakers, we put ourselves in danger of attributing to the child properties it does not yet possess.

Despite these difficulties headway has been made, and a view of the child as something other than a preformed adult is beginning to emerge (see especially Menn, 1978a, 1978b, 1980, 1983; Menyuk & Menn, 1979). A striking aspect of this view, though not, I think, a surprising one, is the lavish variability of the child's productions. In these last few paragraphs, I will briefly consider how we might approach this variability.

Variability within a child. Ferguson (this volume) presents compelling arguments for regarding the word as the unit of contrast in early speech; he defines a word as "...any apparently conventionalized sound-meaning pair." The definition is important, because it draws attention to the fact that a word is not simply a pattern of sound, but a pattern of sound appropriate to a particular situation (Menyuk & Menn, 1979). To discriminate one word from another, to recognize a word and to use it correctly, therefore entail discriminating and recognizing various non-linguistic properties of a situa-

tion. Thus, a child's failure to discriminate or recognize a word in a perceptual test may reflect non-linguistic as much as linguistic factors. Moreover, many of the child's spoken variations may reflect variability in the situations in which the child has heard the word and in the varying salience of its phonetic properties in those situations: the same adult word may then be a different word to the child in different situations.

Nonetheless, highly variable productions of a given word do occur within essentially the same situation. Ferguson (this volume; Ferguson & Farwell, 1975, p. 423, footnote 8) lists ten different attempts by a child (K at approximately 1 year, 3 months) to say pen within one half-hour session. Ferguson comments: "She seemed to be trying to sort out features of nasality, bilabial closure, alveolar closure, and voicelessness." Waterson (1971) describes numerous such instances for her child, P, in similar phonetic terms, noting as a common occurrence that "features" lose their order and become recombined into patterns quite unlike the adult model. Perhaps, however, we would do well to avoid featural terminology. We might attempt a more direct articulatory description as do Menyuk and Menn (1979), describing protowords of one of Menn's (1978a) subjects, Jacob: "...Jacob was varying the timing of front-back articulations against the timing of lowering and raising the tongue" (p. 61). Of course, this is little more than a gloss on phonetic transcriptions. Yet, in the absence of cineradiographic or even acoustic records, the gloss may "...help us see more clearly what it is the child needs to learn and to look at it in a way less coloured by our knowledge of mature linguistic behavior" (Menyuk & Menn, 1979, p. 61; cf. Kent, in press). For we then see the speaking of a word not as a bundling of features into concatenated segments, but as a distribution of interleaved movements of articulators over time (Browman & Goldstein, ms.). In the adult, repeated coordination of particular movements in recurrent patterns has crystallized into structures that form the phonological elements of the language. For the child the movements have yet to be organized.

Here three points deserve emphasis. First, despite the variability of a child's productions, they also display surprising accuracy. The phone classes of Ferguson and Farwell (1975) show much variability in voicing and manner--due perhaps to unskilled timing of closure and release--yet remarkable homogeneity in place of articulation. Also, K's attempts at pen did not include, for example, [gʌk]: Almost every attempt included some recognizable property of the adult word. This means that the acoustic structure of adult words specifies for the child at least some rough pattern of configurations of the vocal tract--necessarily the product of a specialized perceptuomotor link. Yet, second, the link is not precisely predetermined: it must develop. Not only the movements, but their relative timing and sequencing must develop. These are complex processes that almost certainly require active movement for their neural control structures to take form. Perhaps, indeed, it is the normal function of babbling to promote growth of these structures in the left hemisphere. In any event, we are now led to see, and this is my third point, that genetically programmed variability is a condition of the child's learning to speak. In general, the longer the life span of an animal, the longer the period of parental care, and the more complex the mature behavior, the more likely is the behavior to develop through an open genetic program (Mayr, 1974) (though, for an exception, see below). Such a program relies on experience to select and, if necessary, shape the needed behavior from a reservoir of variable responses (cf. Fowler & Turvey, 1978).

Variability among children. As earlier noted, some individual differences in the course of development are genetic or congenital in origin. MacKain (in press) describes several extreme cases of children born without a tongue who approach a surprisingly normal phonetic repertoire by an idiosyncratic path of development. Yet other differences arise from the plasticity of an open system, sensitive to environmental contingencies and equipped with a variable repertoire of responses. Adaptive response to some particular, short-term aspect of the environment may lead an individual down an idiosyncratic path, because the precise order in which the parts of the system assemble themselves is not preordained. Here we may draw a useful analogy with the self-stabilizing processes in embryological development termed "canalization" (Waddington, 1966, p. 48). Waddington describes how various regions of an embryo differentiate into eyes, arms, legs, and so on. Each region has many possible paths to the same end. The exact path is determined, in part, by chance factors in the embryonic environment; equifinality is assured by fixed constraints inside and outside the developing region. Similarly, we may suppose, no single path is prescribed for the development of a phonological system. Many paths, determined by partially fixed, partially variable perceptual, motoric, and social conditions lead to the same end (cf. Lindblom, MacNeilage, & Studdert-Kennedy, 1983).

Certainly, there may be a "normal" path, the product of articulatory proclivity (or "ease") (Locke, 1983) and perceptual salience. But a child can readily be diverted from the path by accidents of the speech it hears or of its physical structure and growth. For example, if final fricatives become salient for a particular child, due to chances of adult lexicon in some recurrent situation, the child may try them and be successful, yet be unable (through lack of consonant harmony in the target word or other "output constraints" [Menn, 1978b]) to execute the initial consonants of the words. A vowel-fricative routine is then established that the child can bring to bear on words that most children would attempt with the standard stop-vowel sequence, followed by a "deleted" fricative (e.g., Waterson, 1971, p. 185). Yet the deviant child will ultimately come upon the same phonological system as its peers.

Here we should note that even quite simple behaviors in non-human animals may develop through an open genetic program. The filial and sexual imprinting of mallard ducklings or domestic chicks on slow-moving objects (such as a walking human, or even a red plastic cube revolving on the arm of a rotary motor [Vidal, 1976]) is well known. The effect is possible because genetic "instructions" are loose: they do not specify the form and color of the mother bird, but only her typical rate of movement. Evolution can afford such imprecision because the normal environment provides the duckling with only one slow-moving object, its mother. If the combination of gross genetic "instructions" and a more or less invariant environment permits essential functions (here, protection from predators and species identification) to develop, there will be no selective pressure for more exact genetic specification.

For the imprinting of precocial birds, the behavior is roughly fixed, while eliciting conditions are only loosely specified. For the development of language, both the behavior and the eliciting conditions are loosely specified.[5] Presumably, the infant has certain minimal, perhaps quite general, capacities (its "initial state"), including sensitivity to the contingencies of its own behavior, the basis perhaps of social responsiveness (Watson, 1972, 1981), while the social environment normally offers the infant certain

more-or-less invariant invitations to interact. So, within weeks of birth we find the infant watching intently its mother's eyes, face and hands, as she talks and plays, and we detect certain inchoate communication patterns in postures of the infant's head, face, and limbs, and in "pre-speech" movements of tongue and lips (Trevarthen, 1979). But at this stage, not even the modality of language is fixed. For if the infant is born deaf, it will learn to sign no less readily than its hearing peer learns to speak. Thus, the neural substrate is also shaped by environmental contingencies; and the left hemisphere, despite its predisposition for speech, is then usurped by sign (cf. Neville, 1980; Neville, Kutas, & Schmidt, 1982; Studdert-Kennedy, 1983, pp. 175 ff. and pp. 219 ff.). In fact, recent studies of "aphasia" in native American Sign Language signers show remarkable parallels in forms of breakdown between signers and speakers with similar left hemisphere lesions (Bellugi, Poizner, & Klima, 1983).

The differences between deaf and hearing individuals are certainly gross. Yet every child grows in its peculiar niche with its peculiar anatomical and physiological biases, and must therefore discover its own "strategy" for fulfilling the human communicative function. (The term "strategy" should be stripped of its cognitive, not to say military, connotations in this context, as it is in standard ethological usage.) Indeed, language, as a sociobiological system, exploits the potential for diverse strategies to mark social groups by channeling speakers into distinctive linguistic styles and dialects--to which, of course, children are highly sensitive (e.g., Local, 1983). Thus, individual differences and individual adaptive response make language a force for social cohesion and differentiation. (For examples of stable diversity within species of bee, treefrog, anemonefish, ruff, and other animals, see Krebs and Davies, 1981, chap. 8).

Finally, individual differences offer an opening for research. Presumably, there are limits on possible strategies. But what these limits may be we do not know. As data from longitudinal studies of individual children accumulate, strategies may cluster, until it is possible to sketch their limits. Such work may lead toward clearer notions of "perceptual salience" and "ease of articulation." Thus, we come back to the constraints on individuals by which phonological elements emerge and phonological systems organize themselves (Lindblom et al., 1983).

References

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology (Vol. 2, pp. 67-96). New York: Academic Press.

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory development and speech perception in infancy. In M. M. Haith & J.J. Campos (Eds.), Infancy and the biology of development (Vol. II, Carmichael's Manual of child psychology, 4th ed.). New York: Wiley and Sons.

Barton, D. (1980). Phonemic perception in children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology (Vol. 2, pp. 97-116). New York: Academic Press.

Baru, A. V. (1975). Discrimination of synthesized vowels [a] and [i] with varying parameters (fundamental frequency, intensity, duration, and number of formants) in dog. In G. Fant & M. A. A. Tatham (Eds.), Auditory analysis and perception of speech (pp. 91-101). New York: Academic Press.

Bellugi, U., Poizner, H., & Klima, E. S. (1983). Brain organization for language: clues from sign aphasia. Human Neurobiology, 2.

Best, C. T., Hoffman, H., & Glanville, B. B. (1982). Development of infant ear asymmetries for speech and music. Perception & Psychophysics, 31, 75-85.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception & Psychophysics, 29, 191-211.

Best, C. T., & Studdert-Kennedy, M. (1983). Discovering phonetic coherence in acoustic patterns. In A. Cohen & M. P. R. v. d. Broecke (Eds.) Abstracts of the Tenth International Congress of Phonetic Sciences (p. 623). Dordrecht: Foris.

Blumstein, S. E. (1981). Phonological aspects of aphasia. In M. T. Sarno (Ed.), Acquired aphasia (pp. 129-155). New York: Academic Press.

Borchgrevink, H. M. (1983). Mechanisms of speech and musical sound perception. In R. Carlson & B. Granstrom (Eds.), The representation of speech in the peripheral auditory system (pp. 251-258). New York: Elsevier Biomedical Press.

Browman, C. P. (1978). Tips of the tongue and slips of the ear: Implications for language processing. UCLA Working Papers in Phonetics 42.

Browman, C. P., & Goldstein, L. Towards an articulatory phonology. Manuscript in preparation.

Burdick, C. K., & Miller, J. D. (1975). Speech perception by the chinchilla: Discrimination of sustained /a/ and /i/. Journal of the Acoustical Society of America, 58, 415-427.

Campbell, R., & Dodd, B. (1979). Hearing by eye. Quarterly Journal of Experimental Psychology, 32, 85-99.

Cowan, N., Leavitt, L. A., Massaro, D. W., & Kent, R. D. (1982). A fluent backward talker. Journal of Speech and Hearing Research, 25, 48-53.

Crowder, R. G. (1983). The purity of auditory memory. Philosophical Transactions of the Royal Society of London, B302, 251-265.

Dewson, J. H. (1964). Speech sound discrimination by cats. Science, 144, 555-556.

Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. Perception & Psychophysics, 16, 513-521.

Eimas, P. D. (1975). Auditory and phonetic coding of the cues for speech: Discrimination of the r-l distinction by young infants. Perception & Psychophysics, 18, 341-347.

Eimas, P. D. (1982). Speech perception: A view of the initial state and perceptual mechanisms. In J. Mehler, E. C. T. Walker, & M. Garret (Eds.), Perspectives on mental representation (pp. 339-360). Hillsdale, NJ: Erlbaum.

Eimas, P. D., & Miller, J. L. (1980a). Discrimination of the information for manner of articulation by young infants. Infant Behavior and Development 3, 367-375.

Eimas, P. D., & Miller, J. L. (1980b). Contextual effects in infant speech perception. Science, 209, 1140-1141.

Eimas, P., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in early infancy. Science, 171, 304-306.

Ferguson, D. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. Language, 51, 419-439.

Fowler, C. A., & Turvey, M. T. (1978). Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. Stelmach (Ed.), Information processing in motor control and learning (pp. 1-40). New York: Academic Press.

Garnica, O. K. (1971). The development of the perception of phonemic differences in initial consonants by English-speaking children: A pilot study. Papers and Reports on Child Language Development (Dept. of Linguistics, Stanford University), 3, 1-29.

Gerstman, L. J. (1968). Classification of self-normalized vowels. IEEE Transactions on Audio- and Electroacoustics, AU-16, 78-80.

Hillenbrand, J. (1983). Perceptual organization of speech sounds by infants. Journal of Speech and Hearing Research, 26, 268-282.

Holmberg, T. L., Morgan, K. A., & Kuhl, P. K. (1977). Speech perception in early infancy: Discrimination of fricative consonants. Paper presented at the 94th Meeting of the Acoustical Society of America, Miami Beach, FL, December 12-16.

Immelmann, K., Barlow, G. W., Petrinovich, L., & Main, M. (Eds.). (1981). Behavioral development. New York: Cambridge University Press.

Jakobson, R. (1968). Child language, aphasia, and phonological universals. The Hague: Mouton.

Jusczyk, P. (1982). Auditory versus phonetic coding of speech signals during infancy. In J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), Perspectives on mental representation (pp. 361-387). Hillsdale, NJ: Erlbaum.

Katz, J., & Jusczyk, P. W. (1980, April). Do 6-month-olds have perceptual constancy for phonetic segments? Paper presented at the International Conference on Infant Studies, New Haven, CT.

Kent, R. D. (1980). Articulatory and acoustic perspectives on speech development. In A. P. Reilly (Ed.), The communication game: Perspectives on the development of speech, language, and non-verbal communication skills. Skillman, NJ: Johnson and Johnson Baby Products Company Pediatric Round Table Series.

Kent, R. D. (in press). The psychobiology of speech development: Co-emergence of language and a movement system. American Journal of Physiology: Regulatory, Integrative and Comparative Physiology.

Kimura, D. (1961). Cerebral dominance and the perception of verbal stimuli. Canadian Journal of Psychology, 15, 166-171.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. Cortex, 8, 163-178.

Kinsbourne, M. (1972). Eye and head turning indicates cerebral lateralization. Science, 176, 539-541.

Krebs, J. R., & Davies, N. B. (1981). An introduction to behavioural ecology. Sunderland, MA: Sinauer Associates.

Kroodsma, D. E. (1981). Ontogeny of bird song. In K. Immelman, G. B. Barlow, L. Petrinovich, & M. Main (Eds.), Behavioral development (pp. 518-532). New York: Cambridge University Press.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America, 66, 1668-1679.

Kuhl, P. K. (1980). Perceptual constancy for speech-sound categories in early infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology (Vol. 2, pp.41-66). New York: Academic Press.

Kuhl, P. K. (1981). Discrimination of speech by non-human animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. Journal of the Acoustical Society of America, 70, 340-349.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. Science, 218, 1138-1144.

Kuhl, P. K., & Miller, J. D. (1978). Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. Journal of the Acoustical Society of America, 63, 905-917.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. Journal of the Acoustical Society of America, 73, 1003-1010.

Lempert, H., & Kinsbourne, M. (1982). Effect of laterality of orientation on verbal memory. Neuropsychologia, 20, 211-214.

Lenneberg, E. H. (1967). Biological foundations of language. New York: Wiley.

Levy, J. (1974). Psychobiological implications of bilateral asymmetry. In S. J. Dimond & J. G. Beaumont (Eds.), Hemisphere function in the human brain (pp. 121-183). London: Elek.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: Evidence for a phonetic mode. Perception & Psychophysics, 30, 133-141.

Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), The production of speech (pp. 217-245). New York: Springer-Verlag.

Lindblom, B. (1984). Can the models of evolutionary biology be applied to phonetic problems? In M. P. R. v. d. Broecke & A. Cohen (Eds.), Proceedings of the Tenth International Congress of Phonetic Sciences (pp. 67-81). Dordrecht: Foris.

Lindblom, B., MacNeilage, P. F., & Studdert-Kennedy, M. (1983). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & Ö. Dahl (Eds.), Explanations of linguistic universals. The Hague: Mouton.

Local, J. (1983). How many vowels in a vowel? Journal of Child Language, 10, 449-453.

Locke, J. (1983). Phonological acquisition and change. New York: Academic Press.

MacKain, K. S. (1982). Assessing the role of experience on infant speech discrimination. Journal of Child Language, 9, 527-542.

MacKain, K. S. (in press). Speaking without a tongue. National Student Language and Hearing Association Journal.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. Applied Psycholinguistics, 2, 369-390.

MacKain, K. S., Studdert-Kennedy, M., Spieker, S., & Stern, D. (1983). Infant intermodal speech perception is a left hemisphere function. Science, 219, 1347-1349.

MacNeilage, P. F. (1983, October). Planning and production of speech. Fourth Silverman Lecture, presented at the Silverman Seminar: Planning and production of speech by normally hearing and deaf people. Central Institute for the Deaf, St. Louis, MO.

MacNeilage, P. F., Hutchinson, J., & Lasater, S. (1981). The production of speech: Development and dissolution of motoric and premotoric processes. In J. Long & A. Baddeley (Eds.), Attention and performance IX (pp. 503-519). Hillsdale, NJ: Erlbaum.

MacNeilage, P. F., Studdert-Kennedy, M., & Lindblom, B. (1984). Primate handedness reconsidered. Unpublished manuscript.

MacNeilage, P. F., Studdert-Kennedy, M., & Lindblom, B. (in press). Functional precursors to language and its lateralization. American Journal of Physiology: Regulatory, Integrative and Comparative Physiology.

Mann, V. A., & Liberman, A. M. (1983). Some differences between phonetic and auditory modes of perception. Cognition, 14, 211-235.

Mayr, E. (1974). Behavior programs and evolutionary strategies. American Scientist, 62, 650-659.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

Mehler, J., Barrière, M., & Jasik-Gerschenfeld, D. (1976). La reconnaissance de la voix maternelle par le nourrisson. La Recherche, 70, 787-788.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. Science, 198, 175-178.

Menn, L. (1978a). Pattern, control, and contrast in beginning speech: A case study in the development of word form and word function. Bloomington: Indiana University Linguistics Club.

Menn, L. (1978b). Phonological units in beginning speech. In A. Bell & J. B. Hooper (Eds.), Syllables and segments (pp. 157-171). Amsterdam: North-Holland.

Menn, L. (1980). Phonological theory and child phonology. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology (Vol. 1, pp. 23-41). New York: Academic Press.

Menn, L. (1983, March). Language acquisition, aphasia and phonotactic universals. Paper presented at 12th Annual University of Wisconsin-Milwaukee Linguistics Symposium.

Menyuk, P., & Menn, L. (1979). Early strategies for the perception and production of words and sounds. In P. Fletcher & M. Garman (Eds.), Language acquisition (pp. 49-70). Cambridge: Cambridge University Press.

Miller, G. A. (1977). Spontaneous apprentices. New York: The Seabury Press.

Miller, J. L., & Eimas, P. D. (1979). Organization in infant speech perception. Canadian Journal of Psychology, 33, 353-367.

Milner, B. (1974). Hemispheric specialization: Scope and limitations. In F. O. Schmidt & F. G. Worden (Eds.), The neurosciences: Third study program. Cambridge, MA: MIT Press.

Milner, B., Branch, C., & Rasmussen, T. (1964). Observations on cerebral dominance. In A. V. S. DeReuck & M. O'Connor (Eds.), Disorders of language (pp. 200-214). Boston: Little, Brown.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception & Psychophysics, 18, 331-340.

Molfese, D. L. (1977). Infant cerebral asymmetry. In S. J. Segalowitz & F. A. Gruber (Eds.), Language development and neurological theory. New York: Academic Press.

Molfese, D. L., Freeman, R. B., & Palermo, D. S. (1975). The ontogeny of brain lateralization for speech and nonspeech stimuli. Brain and Language, 2, 356-368.

Moskowitz, A. I. (1973). The acquisition of phonology and syntax. In I. Hintikka et al. (Eds.), Approaches to natural language. Dordrecht: Reidel.

Neville, H. J. (1980). Event-related potentials in neuropsychological studies of language. Brain and Language, 11, 300-318.

Neville, H. J., Kutas, M., & Schmidt, A. (1982). Event-related potential studies of cerebral specialization during reading. Brain and Language, 16, 316-337.

Ohala, J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.), The production of speech (pp. 189-216). New York: Springer-Verlag.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology (Vol. 1, pp. 93-112). New York: Academic Press.

Oller, D. K., & Eilers, R. E. (1983). Speech identification in Spanish- and English-learning two-year-olds. Journal of Speech and Hearing Research, 26, 50-53.

Oller, D. K., & MacNeilage, P. F. (1983). Development of speech production: Perspectives from natural and perturbed speech. In P. F. MacNeilage (Ed.), The production of speech (pp. 91-108). New York: Springer-Verlag.

Oller, D. K., Wieman, L. A., Doyle, W., & Ross, C. (1975). Infant babbling and speech. Journal of Child Language, 3, 1-11.

Schwab, E. C. (1981). Auditory and phonetic processing for tone analogs of speech. Unpublished doctoral dissertation, State University of New York at Buffalo.

Segalowitz, S. J., & Chapman, J S. (1980). Cerebral asymmetry for speech in neonates: A behavioral measure. Brain and Language, 9, 281-288.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), The production of speech (pp. 109-136). New York: Springer-Verlag.

Studdert-Kennedy, M. (Ed.). (1983). Psychobiology of language. Cambridge, MA: MIT Press.

Studdert-Kennedy, M., & Shankweiler, D. P. (1970). Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 48, 579-594.

Summerfield, Q. (1979). Use of visual information for phonetic perception. Phonetica, 36, 314-331.

Thelen, E. (1983). Learning to walk is still an "old" problem: A reply to Zelazo. Journal of Motor Behavior, 2, 139-161.

Trevarthen, C. (1979). Communication and cooperation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), Before speech (pp. 321-347). New York: Cambridge University Press.

Vidal, J. M. (1976). Empreinte filiale et sexuelle - reflexions sur le processus d'attachement d'apres une etude experimentale sur le coq domestique. Docteur des Sciences These, University of Rennes, France.

Vihman, M. M., Macken, M. A., Miller, R., Simmons, H., & Miller, J. (in press). From babbling to speech: A reassessment of the continuity issue.

Waddington, C. H. (1966). Principles of development and differentiation. London: MacMillan.

Warfield, D., Ruben, R. J., & Glackin, R. (1966). Word discrimination in cats. Journal of Auditory Research, 6, 97-119.

Waterson, N. (1971). Child phonology: A prosodic view. Journal of Linguistics, 7, 179-211.

Watson, J. S. (1972). Smiling, cooing and "The Game." Merrill-Palmer Quarterly, 18, 323-339.

Watson, J. S. (1981). Contingency experience in behavioral development. In K. Immelmann, G. B. Barlow, L. Petrinovich, & M. Main (Eds.), Behavioral development (pp. 83-89). New York: Cambridge University Press.

Werker, J. F. (1982). The development of cross-language speech perception: The effect of age, experience and context on perceptual organization. Unpublished doctoral dissertation, University of British Columbia, Vancouver, B.C.

Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. Child Development, 52, 349-355.

Woodworth, R. S. (1938). Experimental psychology. New York: Holt.

Zaidel, E. (1976). Language dichotic listening and the disconnected hemispheres. In D. O. Walter, L. Rogers, & J. M. Finzi-Fried (Eds.), Conference on Human Brain Function. Los Angeles: Brain Information Service/BRI Publications Office, UCLA.

Zaidel, E. (1978). Lexical organization in the right hemisphere. In P. A. Buser & A. Rougeul-Buser (Eds.), Cerebral correlates of conscious experience (pp. 177-197). Amsterdam: Elsevier/North Holland Biomedical Press.

Footnotes

[1]The periods used here are not fixed "stages" of development. They are simply convenient headings that correspond roughly to a period before babbling (0-6 months) on which most of the infant perceptual research has focussed, and a period of babbling (7-12 months) on which there has been very little perceptual research.

[2]This interpretation assumes that arbitrary groups were, in fact, more difficult to discriminate than "phonetic" groups. Perhaps it is easier to detect a difference between groups, if all members of one group differ from all members of another group on the same dimension ("phonetic") than if each member of one group differs from each member of another on a different dimension (arbitrary). The difference in task difficulty might then be great enough to show up, if the criterial response is itself relatively difficult (head turning), but not if the response is relatively easy (high amplitude sucking).

[3]Jusczyk (1982) makes the same point, proposing the "...possibility...that...recognition of phonetic identities is not achieved until the child is engaged in learning how to read" (p. 365, footnote 3). If "recognition" here means "metalinguistic awareness," Jusczyk may be right. But functional categories surely predate the alphabet, both ontogenetically and historically. The alphabet (like dance notation) can only succeed because its units correspond to functional units of perceptuomotor control. The task for the child, learning to read, is to discover these units in its own behavior.

[4]I am not proposing that language can take any arbitrary form. On the contrary, its general form, that is, its two-leveled hierarchical structure of phonology and syntax, emerges necessarily from its function. Innumerable details of form within these levels must result from more-or-less invariant perceptuomotor, cognitive and pragmatic constraints, of which we know, at present, very little.

INVARIANCE: FUNCTIONAL OR DESCRIPTIVE?*

A comment on C. A. Ferguson's "Discovering sound units and constructing sound systems: It's child's play"

Michael Studdert-Kennedy†

The variability discussed by Ferguson is, of course, quite different from the variability that has been the focus of much speech research since its inception, and especially of research by Ken Stevens. For this focus has been on what we might call lawful variability: the goal has been to discover the invariants presumed to underlie regular variations in the articulatory and acoustic structure of phonetic elements as a function of stress, rate, and context. Ferguson's concern, on the other hand, is with the seemingly unlawful (certainly unpredictable and therefore, in effect, random) variability of early child speech, both within and across children. Moreover, Ferguson's work is mainly concerned with production, while Stevens' interests (at least as they bear on child phonology) have largely been in the problem that acoustic variability poses for perception. Finally, even the unit of variation that occupies Ferguson, namely the word, differs from the familiar units of concern in speech research. In spite (or because) of these differences, I believe that the work Ferguson discusses may carry the seeds of a new and fruitful approach to the notorious puzzles of segmentation and invariance.

My purpose here is to trace some implications of what Ferguson describes, as he follows the emergence of the child's first words over roughly the third half-year of life. The unit of contrast at this stage, Ferguson tells us, is the word defined as "...any apparently conventionalized sound-meaning pair." The emphasis on function is important. The word is a unit of contrast because it is a unit of meaning, offered by the surrounding language and commensurate with the child's cognitive grasp. This does not imply that other structures are not already being put to contrastive use; for they certainly are, as Menn's (1978) study of early intonation, for example, has shown us. However, it is Ferguson's hypothesis that the word is the simplest non-prosodic unit with which a child can begin to accomplish some part of its communicative intent.

An important implication of the claim that the word is the unit of contrast is that smaller units, that is, phone-sized segments and features, are not. This does not mean that acoustic correlates of phones and features can-

---

*To appear in J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: Erlbaum, in press.

†Also Queens College and Graduate Center, City University of New York

Acknowledgment. Preparation of this comment was supported in part by NICHD Grant HD-01994 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

not be described in the utterances of a child. Nor (as we shall see shortly) does it mean that words are perceived by the child as unanalyzed integers. All that it means is that these smaller units have not yet taken on, for the child, the systemic function of contrast that they serve in the adult.

To elaborate these notions somewhat, let us speculate briefly on how the child perceives and produces words. Most early words are open monosyllables, or reduplicated syllables, formed by the child's closing and then opening its mouth, usually while its vocal cords vibrate. What must the child do, if it is to close and open its mouth in such a way that the acoustic consequences will count as a word? (Here I disregard so-called "proto-words," recurrent phonetic structures that cannot be traced to an adult model.) First, of course, the child must execute the act in some appropriate set of circumstances--a remarkable cognitive achievement that we will set aside. Second, from a phonetic point of view, the child must find, in the acoustic structure of an adult word, information that will specify its own articulators' movements (cf. Browman & Goldstein, ms.). Third, the child must execute those movements.

At the risk of laboring the obvious, let us roughly spell the process out. Suppose, for example, that a child utters [mɛ], while reaching for a cup, and that an observing adult happily recognizes an attempt at [mIlk]. Evidently, the acoustic structure of the adult word specified at least the following gestures in a more or less precise temporal arrangement: (1) set larynx into vibration, (2) raise jaw and close lips, (3) lower jaw and open lips, (4) raise velum, (5) raise tongue. Thus, the perceptual representation that controls the child's movements must already have been "segmented" to the extent that it specified the actions of distinct and partially independent articulators.

We may view these actions and their acoustic specifications as precursors of systematic phonetic features, if we wish. But we should not be misled thereby into assuming that the child classifies speech sounds perceptually according to invariant properties shared across contexts. Indeed, evidence for this capacity in infants is quite equivocal (for discussion, see Studdert-Kennedy, this volume).

Consider, here, the ideal case of a child's first word, or, perhaps, first imitation of an adult segmental sound pattern. If the event follows the model sketched above for [mɛ], the child has no need to have "recognized" that components of the acoustic information belong to classes of components whose members occur in other contexts. All that is required is that the acoustic information specify a pattern of articulator action in this word. Thus, for the child, its first word (and indeed every word in its early repertoire) is phonetically unlike every other word in almost every respect. This is the implication, it seems to me, of the claim that the word is the unit of contrast.

To elaborate, let us take the syllables [dae] and [di], treating them, for present purposes, as items in a child's repertoire. The first syllable of the adult models may have had flat or falling, the second rising second and third formant transitions, a frequently cited example of a lack of invariance. However, on the present view, we need not suppose that the perceptual representations controlling the syllable onsets, when the child combines them to utter [daedi], are identical. Rather, if the child is tracking the ges-

tures in the speech it hears, it will find a slightly retracted alveolar contact followed by backward movement of the tongue, in the first syllable and a slightly fronted contact, followed by forward movement of the tongue, in the second, and so will produce just the so-called "coarticulated" pattern it has heard. As the range of contexts in which a child hears and produces alveolar closure and release widens, an auditory-articulatory class may be formed. However, the class qua class initially has no function. Any particular instance of alveolar closure and release is perceived or produced as an idiosyncratic articulatory routine contributing to formation of the particular word to which it belongs.

I will not speculate further on the processes by which recurrent articulatory routines or gestures may crystallize into classes of control structures, or phonemes, contrasting systematically in terms of their defining features. These are matters for the child phonologist. But I have two brief disclaimers.

First, the notions sketched above in no way cast doubt on possible functions of features and phonemes in later language. The function of the phoneme, for example, as a control structure in speaking, is demonstrated by the fact that most normal children can learn to consult their own productions and to write alphabetically (sometimes even before they can read). A system of behavioral notation (as in the alphabet, music, and dance) could only serve as a set of instructions to behave, if the instructions matched already existing control structures. Just as the bicycle was a technological discovery of new behaviors implicit in the cyclical mode of human locomotion, so the alphabet must have been a discovery of new behaviors, reading and writing, implicit in the motor control of human speech.

My second disclaimer is that the view taken here has any bearing on whether we may or may not be able to arrive at satisfactory descriptions of invariant classes in the articulatory and acoustic structures of speech. My intent is merely to raise the possibility that such invariants would be simply descriptive, an outcome rather than a condition of development. Invariants, as invariants, may have no necessary function for the child learning to speak.

References

Browman, C. P., & Goldstein, L. Towards an articulatory phonology. Manuscript in preparation.

Menn, L. (1978). Pattern, control, and contrast in beginning speech: A case study in the development of word form and function. Bloomington, IN: Indiana University Linguistics Club.

Studdert-Kennedy, M. (1984). Sources of variability in early speech development. Haskins Laboratories Status Report on Speech Research, SR-77/78, this volume. Also in J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: Erlbaum, in press.

BRIEF COMMENTS ON INVARIANCE IN PHONETIC PERCEPTION*

A. M. Liberman†

According to the instructions of my hosts, I have ten minutes to tell how I see the matter of invariance. So, getting right to the point, I should say that my concern is with invariance only in the conversion from sound to phonetic structure, then move immediately to the facts that such invariance ought, in my view, to take into account.

> Because of the way we speak, the acoustic information for a phonetic segment commonly comprises a large number and wide variety of cues, most of them dynamic in form. These cues span a considerable stretch of sound, grossly overlap the cues for other segments, and are subject to a considerable amount of context-conditioned variation.
>
> The phonetic perceiving system is sensitive--one might say exquisitely sensitive--to all the acoustic cues. None of them is truly necessary; all are normally used; and their relative importance bears little relation to their salience as it might be reckoned on a purely auditory basis.
>
> Perception of phonetic structure is immediate in the sense that there is no conscious mediation by, or translation from, an auditory base. This is to say, most generally, that listeners are only aware of the coherent phonetic structure that the cues convey, not of the quite different auditory appearances the cues might be expected to have, given their overlap, context-conditioned variation, number, diversity, and dynamic nature. Thus, taking stop consonants and their dynamic formant-transition cues as a particular example, I note that listeners are not aware of the transitions as pitch glides (or chirps) and also as (support for) a stop consonant; listeners are only aware of the stop. Yet these same formant transitions are perceived as pitch glides (or chirps) when--on the nonspeech side of a duplex percept, for example--they do not figure in perception of a phonetic segment.

These facts have two implications relevant to our concern. One is that the invariance between sound and phonetic structure should be sought in a general relation between the two that is systematic but special, not in particu-

---

*Also to appear in J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: Erlbaum, in press.

†Also Yale University and University of Connecticut.

Acknowledgment. This work was supported in part by NICHD Grant HD-01994.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

lar connections that are occasional and discrete. The relation we seek can be seen to be systematic to the extent it is governed by lawful dependencies among articulatory movements, vocal-tract shapes, and sounds, dependencies that hold for all phonetically relevant behavior, not just for specific and fixed sets of elements. The relation has got to be special because the vocal tract and its organs are special structures that behave, most obviously in coarticulation, in special ways. A second implication is that the special relation between sound and phonetic structure is acted on in perception by a system that is appropriately specialized for the purpose.

If the foregoing assumptions are correct, then the invariance in speech is not unique. Rather it resembles, at least grossly, the kinds of special invariances that are found in many perceptual domains. Accordingly, the system that is specialized for phonetic perception can be seen as one of a class of similarly specialized biological devices. All take advantage of a systematic but special invariance between the "proximal" stimuli and some property of the "distal" object. The result is immediate perception of that which it is most important to perceive--namely, the properties that make it possible to identify the invariant distal object.

Consider, as an example, visual perception of depth as determined by the proximal cue of binocular disparity. There is a general and systematic, yet special, relation between the distal property (relative distance of points in space) and the proximal stimulus (disparity). The relation is general and systematic in that it is governed by the laws of optical geometry and holds for all points (within its range) and for all objects, not just for some. The relation is special because it depends on the special circumstances that we have two eyes, that they are so positioned (and controlled) as to be able to see the same object, and that they are separated by a particular distance. Neurobiological investigation has revealed an anatomically and physiologically coherent system--a biological "module," if you will--that is specialized to process the proximal disparity and relate it to the distal depth. Given that specialization, perception of depth is automatic and immediate: there is no conscious mediation by, or translation from, the double images we would see if, in fact, we were perceiving the proximal disparity as well as the distal property it specifies.

Other perceptual phenomena have the same general characteristics. Auditory localization and the various constancies come immediately to mind, and, if we put aside questions about phenomenal "immediacy," so too do such processes as those that underlie echolocation in bats and song in birds. These are surely specializations if only because each such process, or module, is as different from every other as is the invariant relation it serves. The phonetic module differs from many of the others in at least two ways.

To make one of the differences clear I would turn again to binocular disparity and depth perception as representative of a large class. In this case the distal object is "out there," a physical thing in the narrow sense of physical, and the invariant relation between its properties and those of the proximal stimulus is determined, as already indicated, by optical geometry and the separation of the eyes. In speech, however, the distal object--a phonetic structure--is a physiological thing, a neural process in the talker's brain, and the invariant relation between its properties and those of the proximal sound is determined in large part by neuromuscular processes internal to the talker but available also to the listener. Thus, the specialized phonetic

module might be expected to incorporate a biologically based link between production and perception. Such a link is not part of the disparity module or of the other perceiving modules it exemplifies, though it may very well characterize the "song center" module of certain birds.

A second important difference in the nature of the invariance (and its module) has to do with the question: What turns the module on? In the case of binocular disparity, the answer is a quite specific characteristic of the proximal stimulus--namely, disparity. Notice, however, that disparity has no other utility for the perceiver but to provide information about the distal property, depth. There are, accordingly, no circumstances in which the perceiver could use the proximal disparity as a specification of, or signal for, some other property. This is to say that disparity and the depth it conveys do not compete with other aspects of visual perception such as hue, form, etc., but rather complement them. Not so in phonetic perception. There is, first of all, the fact that the speech frequencies overlap those of nonspeech. More to the point, the formant transitions that we don't want to perceive as chirps when we are listening to speech are very similar to stimuli that we do want to perceive as chirps when we are listening to birds. Thus, almost any single aspect of the proximal stimuli can be used for perception of radically different distal objects: phonetic structures in a talker's head or acoustic events and objects in the outside world. What follows is that the module can hardly be turned on by some specific (acoustic) property of the proximal stimulus. Not surprisingly, then, we find in research on speech perception that the module is, in fact, not turned on that way, but rather by some more global property of the sound. Thus, just as in the perception of phonetic segments all cues are responded to but none is necessary, so too in identifying sound as speech.

How, then, is the module turned on? What invariant property of the sound causes the listener to perceive that the distal object is a phonetic structure and not some nonlinguistic object or event? I offer a suggestion. Suppose that auditory stimuli go everywhere in the nervous system that auditory stimuli can go, including, of course, the language center. Suppose, further, that the language center applies the principle: if the shoe fits, wear it. What is decided, then, by the language center is the answer to the question: could these sounds, taken quite abstractly, have been produced by linguistically significant articulatory maneuvers, also taken quite abstractly? If the answer is yes, then the module takes over the purely phonetic aspects of the percept, and the auditory appearances are inhibited. (Auditory aspects that are irrelevant to the phonetic, such as loudness or hoarseness, are perceived, of course, as attributes of the same distal object.) If the answer is no, then the phonetic module shuts down and the ordinary auditory appearances of the stimuli are perceived. Hence the common experience of those who work with synthetic speech that when the sound includes configurations that the articulatory organs cannot produce, as well as those it can, the percept breaks, correspondingly, into nonspeech and speech. Phenomenally, the nonspeech stands entirely apart from, and bears no apparent relation to, the speech, even though the acoustic bases for these wholly distinct percepts were perfectly continuous. The same arrangement for turning the module on (or off) might account for the fact that certain kinds of acoustic patterns--for example, sine waves in place of formants--can be perceived as speech or as nonspeech depending on circumstances that in no way alter the acoustic structure of the stimulus. It also helps to explain how, as in the unnatural procedures of duplex perception, we can disable the mechanism that forces the choice be-

tween speech and nonspeech, and so create a situation in which exactly the same proximal formant transition is simultaneously perceived (in the same context and by the same brain) as critical support for a stop consonant and also as a nonspeech chirp. At all events, there is a kind of competition between phonetic perception and other ways of perceiving sound. A consequence is that the phonetic module produces a more or less distinct mode of perception in a way that modules like depth perception do not. This phonetic mode accommodates a class of distal objects that are distinguished, not only by their role in language, but also by the special nature of the invariant relation by which they are connected to sound.

PHONETIC CATEGORY BOUNDARIES ARE FLEXIBLE*

Bruno H. Repp and Alvin M. Liberman†

Introduction

In the grammatical domains of language we find no gradients, only categories. Thus, gradations of, for example, tense (present - past), form class (noun - verb), or even word (night - day) are everywhere absent. Indeed, they are impossible, for syntactic, morphologic, and phonologic devices do not permit of continuous variation. At the surface of language, however, the situation is different. There, in the relation between phonetic structure and sound, the role of the segments is categorical--a segment is, for example, [d] or [g], not something in between--but the sound can vary continuously. That being so, at least in synthetic speech, we can ask whether the phonetic segments are categorical, not only in their linguistic function, but also in the way they are perceived. The answer is a qualified "yes." Other things equal, stimuli belonging to the same phonetic category are more difficult to discriminate than stimuli on opposite sides of a phonetic boundary. This phenomenon has long been known as "categorical perception" (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). The research it has generated, which was recently reviewed by one of us (Repp, 1984), is largely concerned with the ability of listeners to detect stimulus differences within the categories--that is, with the degree to which perception is perfectly categorical--and with the conditions under which that ability can be made to vary. Our concern in this chapter is rather with the conditions under which the locations of the categories on a continuum can be shown to vary, and with the implications of that variation for a theory about the nature of the categories. More particularly, we will be concerned with the boundaries between the categories (and with their movement), so before considering the relevance to theory, we should justify our concern with the boundaries.

We take the boundary to be the point along the appropriate (acoustic) stimulus continuum at which subjects classify stimuli into alternative categories with equal probability. In the typical case of two (adjacent) categories, this is simply the point corresponding to the 50-percent cross-over of the response function. If more than one stimulus dimension is varied, category boundaries may be represented by contours in a multidimensional space (see, e.g., Oden & Massaro, 1978). The standard method of obtaining category boundaries is to present a set of stimuli repeatedly (and in random order) for identification as members of one class or another. Several alternative meth-

---

*In Steven N. Harnad (Ed.), Categorical perception. New York: Cambridge University Press, in press.

†Also University of Connecticut and Yale University.

Acknowledgment. Preparation of this book chapter was supported by NICHD Grant HD-01994.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

ods--for example, a method of adjustment--have been used, but all yield similar boundaries (Ganong & Zatorre, 1980).

Why do we take account only of the boundaries? After all, it is the categories themselves, rather than the boundaries between them, that play the important role in speech communication. Why not, then, deal with some appropriate exemplar--the prototype, as it were--of the category? A sufficient reason is that, until recently, no one had used methods designed to identify the prototypes. Worse yet, the application of such methods has so far not yielded entirely satisfactory results (Samuel, 1979, 1982). The measurement of boundaries, on the other hand, has long been common in research on speech, so the data are plentiful. Moreover, the boundaries do inform us about the categories and, under some specifiable conditions, about their positions on the appropriate acoustic continua. And, finally, as we will say below, it is the boundaries, not the prototypes, that are central to the assumptions underlying at least one of the important theories about the categories.

Still, it is important to keep in mind that the location of a category boundary is determined, not only by the listeners' internal representations (the prototypes) of the categories, but also by the criterion they adopt for deciding between two competing categories, which makes the boundary vulnerable to biasing influences of various kinds. In principle, at least, a change in the location of a boundary may result either from a change in one or the other (or both) of the category prototypes, or from a criterion shift.

It is important to know whether, and under what conditions, the boundaries between phonetic categories are flexible, because the question bears on two very different hypotheses about the processes that underlie the categorization. According to one hypothesis, the perceived categories result from psychophysical discontinuities that directly reflect the characteristics of the auditory system. Thus, given an acoustic stimulus continuum appropriate for some phonetic distinction, a category boundary is assumed to fall naturally at a point on the continuum where, owing to the way the ear works, differential sensitivity undergoes a sudden change. Perhaps the most general implication of this hypothesis is that auditory categories are the stuff of which phonetic categories are made. Put another way, the implication is that articulatory gestures are so governed as to produce sounds that fit within the categories that the auditory system happens to provide. Accordingly, we will refer to this as the "auditory" hypothesis. By any name, it is the hypothesis, referred to earlier, that deals directly with the boundaries of the categories rather than their ideal exemplars or prototypes. As for movement of category boundaries, that is allowed under this hypothesis, but only as a result of psychoacoustic factors that apply to auditory perception in general, and only to the extent that such factors can actually modify the patterns of differential sensitivity on which the auditory boundaries rest.

The other hypothesis is that the boundaries are determined by category prototypes that reflect typical productions of the relevant speech segments. Accordingly, the prototypes and the boundaries between them need not conform to discontinuities in the auditory system, but are, instead, free to be precisely as flexible as the acoustic consequences of the articulatory gestures require. In fact, considerable flexibility may be demanded. The efficiency of phonetic communication depends crucially on the ability of the several articulators to produce successive phonetic segments at the same time (or with considerable overlap), and also to accommodate in other ways to changes

in phonetic context and rate. These maneuvers can produce systematic changes in the way a particular phonetic segment is represented in the sound. If the perceiving apparatus were not flexibly responsive to those changes, communication would break down, or so it seems. Moreover, the inventory of phones will itself change as language changes, and this, too, requires flexibility in the prototypes. Our hypothesis is that a link between perception and production (in most general terms) enables the category prototypes to respond appropriately to articulatory or co-articulatory adjustments, and so to mirror the talker's phonetic intent. Needing a convenient name to refer to this hypothesis, and wishing to distinguish it from the "auditory" hypothesis we described first, we will call it "phonetic."[1]

Our aim in this chapter is to bring together the many data that demonstrate flexibility of a kind the phonetic hypothesis leads us to expect. These pertain to the influences on perceived phonetic boundaries of such factors as phonetic context, speaking rate, the mix of acoustic cues, and linguistic experience. But there are other effects on the perceived boundaries about which the auditory and phonetic theories are neutral. These include the consequences of varying the range, frequency, and order of the stimuli, as well as such phenomena as contrast and adaptation. Since effects of that kind need to be distinguished from those that are more directly relevant to the auditory and phonetic theories, we will consider them first. We will note, however, that even these "simple" effects sometimes follow patterns that seem difficult to reconcile with a purely auditory theory, and that suggest that speech-specific perceptual criteria may play a role in certain situations. Our review will be selective and focus especially on these instances.

Stimulus Sequence Effects

Under this heading we consider influences on the perception of speech stimuli exerted by other, similar stimuli preceding or following them in a sequence. These effects need to be distinguished from the "stimulus structure effects" discussed later, which concern perceptual dependencies within a single coherent speech stimulus or influences entirely due to factors within the listener.[2]

It is generally agreed that vowel identification--of isolated steady-state vowels, at least--is highly susceptible to all sorts of stimulus sequence effects. On the other hand, the identification of consonants, and of stop consonants in particular, is more stable and less sensitive to stimulus context. This difference parallels the well-known difference between these two stimulus classes in the extent of "categorical perception"; indeed, the criterion of "absoluteness" (i.e., independence of surrounding stimuli) constituted part of the classical definition of categorical perception (Studdert-Kennedy et al., 1970). "Context sensitivity" in a sequence may be distinguished on logical grounds, however, from the extent of the subject's reliance on category labels in discriminating between stimuli (Lane, 1965; Repp, Healy, & Crowder, 1979), and these two aspects of categorical perception can, to some extent, be dissociated experimentally (Healy & Repp, 1982).

Local Sequential Effects

Local sequential effects--typically, influences of a preceding stimulus on the identification of a following stimulus--may occur in any random test sequence. These effects are pervasive in absolute identification, magnitude

estimation, and other psychophysical tasks involving nonspeech stimuli. Surprisingly, there have been very few attempts to determine the extent of sequential effects in standard speech identification tests, where stimuli are presented in random order. Of course, there is an indirect test in the shape of the labeling function, since it can be steep only if sequential effects are relatively small.

In several studies of speech-sound identification, however, the stimuli have been presented in balanced arrangements specifically designed for the assessment of sequential context effects. In one of the earliest of these studies, Eimas (1963) called for identification of stimuli presented in ABX triads of the sort often used in discrimination tasks, and found large context effects for isolated vowels (see also Fry, Abramson, Eimas, & Liberman, 1962) and smaller, but by no means negligible, effects for both the voicing and place dimensions of stop consonants. All effects were contrastive--that is, a stimulus tended to be classified into a category different from that of the stimulus it was paired with--and the magnitude of the effect increased with the acoustic distance between adjacent stimuli. Comparable results have been obtained more recently by, among others, Healy and Repp (1982).

Although sequential effects are generally considered to be common to speech and nonspeech stimuli, there are some intriguing differences. For example, it has been found in several studies that the magnitude of the contrast effect is greater for continua of isolated vowels than for nonspeech continua such as pitch or duration (Eimas, 1963; Fujisaki & Shigeno, 1979; Healy & Repp, 1982; Shigeno & Fujisaki, 1980). While it is possible that the difference is to be accounted for by the more complex acoustic (and auditory) nature of the vowels (and there are also problems with comparing the magnitudes of contrast effects across different stimulus continua), it may, with equal plausibility, be taken to reflect a flexibility of categorization peculiar to the class of vowel sounds, a class that happens to carry the major burden of dialectal variation and language change.

If two or more stimuli in a sequence must be held in memory before a response is permitted, as in the procedure of Eimas (1963) described above, the effects of the stimuli on each other are retroactive as well as proactive. Interestingly, retroactive effects tend to be larger than proactive effects for isolated vowels, while the opposite tends to be the case for all other types of stimuli examined, whether speech or nonspeech (Diehl, Elman, & McCusker, 1978; Healy & Repp, 1982; Shigeno & Fujisaki, 1980). This finding, like the one having to do with the magnitude of contrast, may be explicable by acoustic stimulus properties alone, or it may reflect a specific tendency, derived perhaps from experience with fluent speech, to revise tentative decisions about vowel categories in the light of later information.

One reason we consider that even simple sequential effects may exhibit speech-specific patterns is that these effects almost certainly take place in two quite distinct ways, one reflecting a sensory effect, the other a judgmental effect (see Simon & Studdert-Kennedy, 1978). That is, there may be an effect of a preceding stimulus on the sensory representation of a following stimulus (as well as the reverse, if both are held in a precategorical memory store), but the judgment of a stimulus may also be affected by the response that was assigned to the preceding or following stimulus, usually in a contrastive fashion. Whereas the purely sensory effects are presumably shared by speech and nonspeech stimuli and are sensitive to factors such as spectral

similarity and temporal proximity (Crowder, 1981, 1982), the special structure and function of phonetic categories may produce criterion shifts in the response domain that are specific to speech. Although a clear separation of stimulus and response effects has rarely been achieved in speech experiments, separate studies provide evidence for each type. Thus, Crowder (1982) has shown that proactive contrast effects for isolated vowels decrease with temporal separation over about 3 s in a manner that parallels the decay of auditory sensory storage in other paradigms. On the other hand, Sawusch and Jusczyk (1981) found that sequential contrast depended more on the perceived category of the preceding stimulus than on its acoustic structure. Judgmental effects may depend in part on whether or not a response to the contextual stimulus is required: A comparison of Crowder's (1982) data with those of Repp et al. (1979) for isolated vowels suggests that proactive contrast effects are reduced when only the second stimulus in a pair requires a response. (It goes almost without saying that retroactive contrast effects would be reduced or eliminated if only the first stimulus in a pair were responded to.)

The distinction between sensory and judgmental components of sequential effects is also familiar in nonspeech psychophysics (e.g., Petzold, 1981) and is compatible with Braida and Durlach's (1972) two-factor theory of perceptual coding (see Macmillan's chapter, this volume). Thus, Petzold (1981) has found that preceding stimuli exert a contrastive effect while preceding responses exert an assimilative effect. On the other hand, Shigeno and Fujisaki (1980) have proposed a two-factor model for sequential effects in speech and nonspeech that predicts precisely the opposite. The limited data available suggest, on the contrary, that for speech both components of sequential effects are contrastive in nature.

Global Sequential (Range-Frequency) Effects

Shifts in phonetic category boundaries may occur as a consequence of variations in the overall composition of a stimulus sequence--that is, the range of stimuli employed and the frequency of occurrence of the individual stimuli. In general, if the stimulus range is shifted or expanded in a certain direction, the boundary will shift in the same direction; and if one stimulus (typically one of the endpoints, the "anchor") occurs more frequently than other stimuli, the boundary will shift toward it. In other words, the effects are contrastive in nature, and, in the case of speech sounds, they exhibit variations in magnitude similar to those observed for simple sequential effects: For stop consonants varying in place or voicing, the effects are small (Brady & Darwin, 1978; Rosen, 1979), while for isolated vowels (Sawusch & Nusbaum, 1979), certain other consonantal contrasts (Repp, 1980), and even for stop consonants in Polish (Keating, Mikos, & Ganong, 1981), they may be quite large.

An interesting asymmetry has been observed in the anchoring paradigm for isolated vowels (Sawusch, Nusbaum, & Schwab, 1980): An analysis of anchoring effects on an /i/-/I/ continuum suggested that the effect of the /i/ anchor was due to sensory adaptation while that of the /I/ anchor represented a change in response criterion. In a recent and similar study, in which the anchor always came first in a stimulus pair and only the second stimulus required a response, Crowder and Repp (1984) found an effect of /i/ but not of /I/. The explanation for this asymmetry may be found in the acoustics of the stimuli; alternatively, it may be owing to the special status of /i/ as one of the corners of the vowel space.

We should note, perhaps, that although range-frequency effects are usually considered to derive from stimulus context beyond the immediate local environment, they are often confounded with sequential probabilities: If a given endpoint stimulus (the anchor) occurs more often than other stimuli, the probability that a given stimulus is immediately preceded by the anchor will be increased relative to an equal-frequency (or a different anchoring) condition. Similarly, if the stimulus range is shifted or expanded in one direction, the likelihood that certain critical stimuli are preceded by other stimuli from that part of the continuum is increased. Therefore, range-frequency effects may in many cases be just local sequential effects in disguise. The extent to which nonlocal stimulus context makes any additional contribution has, to our knowledge, not been ascertained experimentally for speech stimuli. It is possible, however, that the frequent occurrence of a single stimulus has an additional adapting influence not evident in regular balanced stimulus sequences. In that sense, the anchoring paradigm approximates the selective adaptation paradigm, to be discussed next.

Selective Adaptation

In selective adaptation experiments, an adapting stimulus (frequently one or the other endpoint stimulus of a speech continuum) is presented repeatedly many times before responses to a few test stimuli are collected. The original motivation for using this paradigm in speech research was the assumption that the effects of the adapting stimulus might reveal the existence and nature of "phonetic feature detectors" (Eimas & Corbit, 1973; see Remez's chapter, this volume). Apart from the difficulty of conceiving that phonetic features (e.g., place, manner, voicing) could possibly be perceived by detectors that respond to such simple features as the auditory analogs of edges and angles in vision (see, e.g., Diehl, 1981; Studdert-Kennedy, 1981; Remez, this volume), a large number of experiments suggest that the effect of selective adaptation take place primarily at the auditory, not the phonetic (judgmental) level. (However, see Elman, 1979.)

The most striking demonstrations of the auditory (as opposed to the phonetic) nature of selective adaptation were provided in two recent studies. In one of these, Roberts and Summerfield (1981) presented audiovisual adapting stimuli that, due to the overriding influence of a conflicting visual display, were never classified into the category normally associated with the auditory stimulus. Nevertheless, the audiovisual adaptors had exactly the same influence on the identification of auditory test stimuli as did purely auditory adaptors. Thus, the phonetic category assigned to the adaptors seemed to play no role in selective adaptation. A similar result was obtained by Sawusch and Jusczyk (1981), who used adaptors of the form /spa/, in which the stop consonant was phonetically classified as "p" but acoustically identical with the initial "b" in /ba/. The adapting effects of /spa/ and /ba/ did not differ.[3] These studies, together with several earlier attempts to dissociate acoustic and phonetic stimulus properties (Blumstein, Stevens, & Nigro, 1977; Sawusch & Pisoni, 1976), suggest that selective adaptation with speech is an exclusively auditory phenomenon. Even though studies of interaural transfer of adaptation effects suggest more than one site at which adaptation takes place (Ganong, 1978; Sawusch, 1977), both of these sites appear to be auditory (i.e., nonphonetic) in nature.

There are two types of evidence, however, that do indicate some involvement of phonetic processing in selective adaptation. One has to do with the influence of the listeners' native language. The relevant finding is that selective adaptation effects on the same stimulus continuum are different for American and for Thai listeners, as independently demonstrated by Donald (1976) and Foreit (1977). The continuum was one of stop consonants varying in voice onset time (VOT), ranging from prevoiced (voicing lead) to devoiced (0 ms VOT) to aspirated (voicing lag). For American listeners, who do not distinguish prevoiced and devoiced stops, a -60 ms VOT and a 0 ms VOT adaptor had the same effect on the category boundary. For Thai listeners, on the other hand, who have three separate categories on the continuum, only the 0 ms adaptor affected the devoiced-aspirated boundary while the -60 ms adaptor was ineffective. This finding agrees with earlier results of Cooper (1974) showing that, on a place-of-articulation continuum divided into three categories, adapting stimuli affected only the adjacent but not the remote category boundary.

The other piece of evidence for a role of phonetic categorization in selective adaptation comes from studies that have revealed differences in the effectiveness of adaptors as a function of their distance from the category boundary. In general, the effectiveness of an adaptor increases with its distance from the boundary (Ainsworth, 1977; Cole & Cooper, 1977; Miller, 1977a), unless it crosses another phonetic boundary (Cooper, 1974; Donald, 1976; Foreit, 1977). Of course, this may be just another instance of the well-confirmed fact that the spectral similarity of adaptor and test stimuli is the major determinant of the size of the adaptation effect. In other words, the distance effect may have a purely auditory explanation. In a recent study, however, Miller et al. (1983) demonstrated that, even if no other phonetic boundary intervenes, the adaptation effect does not increase indefinitely as the adaptor moves away from the boundary, but instead reaches a maximum and then declines (or, for some subjects, remains on a plateau). The adaptor that produces the maximum effect has characteristics that may reasonably be assumed to be optimal for its category, which led Miller et al. to conjecture that the size of the adaptation effect is related to the adaptor's distance from the listener's internal category prototype. Preliminary support for this hypothesis was obtained by Miller et al. in a condition in which the category boundary on a /ba/-/wa/ continuum, and with it the presumable location of the /wa/ prototype (cf. Miller & Baer, 1983), was made to shift by reducing the duration of the syllables. The peak in the function relating the size of the adaptation effect to the location of the adaptor on the continuum shifted accordingly, as predicted.

Even stronger support for a role of "category goodness" in selective adaptation comes from a study by Samuel (1982). He first asked his subjects to locate the optimal /ga/ on a /ga/-/ka/ VOT continuum. The subjects were then divided into two groups--those with short-VOT and those with long-VOT prototypes. Two adapting stimuli matching the two average prototypes were then selected. For each group of subjects, the adaptor matching the group's prototype produced the larger boundary shift. Since exactly the same adaptors were used for both groups, the listeners' internal category prototype seemed to be responsible for the magnitude of the adaptation obtained.

These recent results lead to the tentative conclusion that selective adaptation takes place at an auditory level that is <u>phonetically relevant</u>. Perhaps this should not come as a surprise. The adapting stimuli, after all,

are speech and therefore are phonetically relevant auditory patterns. Conversely, the internal standards or category prototypes against which listeners presumably compare stimuli in the process of categorization must entail detailed auditory specifications; otherwise, in the absence of a common metric, the comparison would be impossible. Selective adaptation may then be viewed as a temporary modification of the prototype itself--a weakening of the criterial specifications that is proportional to the degree to which the auditory input meets those specifications. With this interpretation, the results reviewed above can be reconciled with the numerous earlier demonstrations of "purely auditory" effects in selective adaptation.

From this vantage point, the various "low-level" effects reviewed so far--sequential contrast, range-frequency effects, and selective adaptation--are relevant to the topic of our paper, the flexibility of phonetic boundaries. In essence, the data seem to show that not even a psychophysical procedure like selective adaptation has its effects exclusively at a "general auditory level" of processing; rather, as long as the adapting stimuli are speech, their effects reflect the extent to which they engage the speech processing apparatus. Since speech stimuli ordinarily engage the mechanisms of phonetic categorization (even in the absence of an overt or covert response), selective adaptation with speech is properly viewed as a speech-specific phenomenon--a modification of the frame of reference within which speech stimuli are interpreted. The same is true for range-frequency and sequential contrast effects, except that overt responses to contextual stimuli may have additional effects at a judgmental level. In other words, although speech must pass through the auditory nerve, there may be no "general auditory" level of representation beyond the peripheral transduction. Speech perception takes place within a pre-established frame of reference, and the auditory representation of speech cannot be separated from the (equally "auditory") internal structures, due to cumulative experience in conjunction with biological predispositions, through which the incoming information is filtered.

Stimulus Structure Effects

Under this heading we consider perceptual dependencies that arise among different components of a single coherent speech stimulus. That stimulus may be as short as a single syllable or as long as a whole sentence. Stimulus structure effects, even though they are most easily revealed in the laboratory, are closer to the real life situation than the stimulus sequence effects discussed in the preceding section, which represent or exploit artifacts of test sequence construction. Although the experimental induction of selective adaptation or sequential contrast may be useful for the purpose of probing perceptual mechanisms, there is no reason to believe that these phenomena (as distinct from the mechanisms they reveal) play any significant role in the perception of coherent speech. The various effects discussed in the present section, on the other hand, have more direct implications for normal speech perception, as they reflect the perceptual functions of integration and normalization that make speech perception so effortless and efficient.

Cue Integration Effects

It is well known that distinctions among phonetic segments rest on a multiplicity of acoustic cues in the speech signal. Typically, these many cues are acoustically diverse, relatively widely distributed in time, and overlapped with cues for other segments. Yet the perceiver somehow integrates

these diverse and distributed aspects of the speech signal to recover the phonetic structure of the message (Liberman & Studdert-Kennedy, 1978; Repp, Liberman, Eccardt, & Pesetsky, 1978). Exactly how the individual acoustic cues are characterized depends to some extent on the methods of analysis and experimental manipulation and on the descriptive framework chosen by the investigator. From a purely acoustic point of view, however, they seem in most cases to be incoherent. From an articulatory point of view, on the other hand, they make sense--that is, they reflect a unitary event in the domain of articulatory planning.[5]

The statement that there are multiple cues for each phonetic contrast must be qualified by the fact that some cues are more important than others. That is, some cues are easily overridden by others. Listeners' sensitivity to the weaker cues can be demonstrated in the laboratory by eliminating the stronger ones or by setting them at ambiguous values. From the existing evidence it can indeed be concluded that, given the opportunity, listeners will make use of any cue for a given phonetic distinction (Bailey & Summerfield, 1980). This general observation suggests that, as Bailey and Summerfield (1980) have pointed out, the concept of cue has limited theoretical relevance. As a practical matter it is useful, even essential, in dealing with the acoustic basis of speech perception. But the sensitivity to the many and various cues for a phonetic segment suggests, as we have already implied, that listeners are perceiving just what all the cues have in common--viz., some economical representation of the coherent process underlying the peripheral articulation.

The relevance of cue integration to the topic of our chapter is evident when we consider that a phonetic category boundary is usually determined on a continuum of stimuli varying in only one important cue dimension. The flexibility of that phonetic boundary may then be assessed by introducing other, usually less important, cues that favor either one or the other response alternative. That boundaries are indeed flexible in this particular sense has been demonstrated in numerous studies. (For a recent review, see Repp, 1982.) By definition, phonetic boundaries are located at the point of maximal ambiguity, where weaker cues have their strongest effect. The perceptual cue integration, or phonetic "trading relation," revealed by the boundary shift generally takes place without the listener's awareness. Perception tends to remain categorical even in the presence of multiple acoustic differences among stimuli (see, e.g., Fitch, Halwes, Erickson, & Liberman, 1980.)

The ubiquity of trading relations among acoustically diverse cues provides one of the strongest arguments against theories that predict fixed boundary locations on any acoustic speech continuum. In many cases, cues are so disparate as to be extremely unlikely to engage in any direct psychoacoustic interaction. Rather, what seems to unite them is that they are common consequences of the articulatory gestures that differentiate phonetic segments; at the same time, they are members of the set of structural acoustic differences that characterize a particular phonetic contrast. To cite only one specific example: The primary cue for the /s/-/ʃ/ distinction is the spectrum of the fricative noise, but a secondary cue is provided by the voiced formant transitions following the noise. The phonetic boundary on an /s/-/ʃ/ continuum, obtained by varying the spectral properties of the fricative noise, is at different locations depending on whether the formant transitions are appropriate for /s/ or for /ʃ/ (Mann & Repp, 1980). Considering that the fricative noise is of relatively long duration, produced by a different

source, and of a spectral composition quite different from that of the following signal, there is little reason to expect any direct effect of the formant transitions on the auditory representation of the fricative noise. Indeed, when listeners are led to focus on the "pitch" of the fricative noise (rather than on the phonetic fricative category), there seems to be no influence of the following formant transitions on their judgments (Repp, 1981). Thus, the perceptual integration of the cues provided by fricative noise spectrum and formant transitions seems to be phonetically motivated and related to the fact that different values of both cues are consistently correlated with different places of fricative production. Similar arguments may be applied to other phonetic trading relations, even including those that could, in principle, result from some psychoacoustic interaction.

Feature Integration Effects

The trading relations discussed in the preceding section (and reviewed by Repp, 1982) take place among cues to a single phonetic feature--e.g., voicing or place of articulation. This is a consequence of the fact that the phonetic categories constituting the endpoints of a speech continuum nearly always differ only in a single feature. Here we consider a related class of effects that reveals perceptual dependencies among cues to different features of the same phonetic segment. The main reason for considering these effects separately is that they give the theorist an additional degree of freedom: Feature interactions may be hypothesized to occur after a process of "feature extraction" but before assembly of the features into a phonetic segment (see, e.g., Miller, 1977b; Sawusch & Pisoni, 1974). For theorists who instead postulate either direct psychoacoustic interactions among the cues or reference to phoneme- or syllable-sized prototypes, the effects considered here are further instances of cue integration (cf. Oden & Massaro, 1978).

The literature on genuine feature integration effects is rather small, for it is difficult to vary cues for different features in a strictly orthogonal fashion. A well-known finding is that the voicing boundary on a VOT continuum is at increasingly larger voicing lags for labial, alveolar, and velar stop consonants (Lisker & Abramson, 1970). In most studies, however, the duration of the first-formant transition, which itself constitutes a voicing cue (as well as a weak cue for place of articulation) covaried with place of articulation, so that the boundary shifts may be considered as being due to a simple trading relation among voicing cues. In one experiment, however, the F1 transition was held constant (with only the F2 and F3 transitions varying to cue differences in place of articulation), and a small but reliable voicing boundary shift as a function of place of articulation was obtained (Miller, 1977b). (See, however, Massaro & Oden, 1980, for a failure to replicate this result.) Subsequently, Miller (1977b) showed that the boundary on a labial-alveolar place of articulation continuum shifted depending on whether the stop consonants were synthesized as nasal, voiced, or voiceless. She interpreted these results as revealing processing dependencies among phonetic features. An alternative interpretation has been proposed in a model that builds feature dependencies into prespecified criterial feature values and so avoids any processing interactions after the feature extraction stage (Massaro & Oden, 1980; Oden & Massaro, 1978). Because of the built-in dependencies, however, the model rests on the assumption of phoneme- or syllable-size prototypes and merely pays lip service to phonetic features.

Feature interactions of the kind observed by Miller (1977b) presumably reflect the inherent nonorthogonality of articulatory features and their acoustic correlates. Clearly, the binary feature matrix devised by phonologists is inadequate from a phonetic viewpoint. Initial velar stops, for example, because of their longer VOTs, simply are relatively "more voiceless" than labial stops. The possibility of psychoacoustic interactions among signal components must be considered, but there is no well-supported psychoacoustic explanation for the observed feature interactions.

One case in which a psychoacoustic interaction between feature dimensions can definitely be ruled out is the finding (Carden, Levitt, Jusczyk, & Walley, 1981) that, given a single continuum of formant transitions, listeners place the phonetic boundary at different locations depending on whether they are instructed to hear the stimuli as stops ([ba], [da]) or as fricatives ([fa], [θa]). This can only be accounted for as an adjustment--and apparently a perfectly automatic one--for the fact that the places of production are somewhat different for the two stops from what they are for the fricatives. Hence, it becomes yet another example of the rule that phonetic categorization is guided by internal criteria that reflect the prototypical acoustic and articulatory characteristics of speech.

Segmental Context Effects

A third class of perceptual interactions taking place within a single utterance concerns perceptual dependencies among cues for different phonetic segments. While the conceptual distinction from the two classes discussed earlier (integration of cues to the same feature, or to different features of the same segment) is straightforward, practical distinctions are somewhat fuzzy because acoustic cues generally cannot be apportioned exclusively to one or the other phonetic segment. However, an experimental dissociation is usually possible between those signal aspects that provide weak (coarticulatory) cues to one segment and those that are strong and sufficient cues for a different segment, even when both very nearly coincide in time.

For example, take the effect of a following vowel on fricative perception, investigated--among others--by Mann and Repp (1980). The periodic signal portion following a fricative noise necessarily has formant transitions characteristic of the fricative's place of production, which contribute to the fricative percept, particularly when the fricative noise spectrum carries little distinctive information (Carden et al., 1981; Mann & Repp, 1980). Therefore, this effect belongs under the heading of cue integration. The identity of the vowel itself, however, is quite independent of the preceding fricative and therefore cannot provide any direct cues to fricative place of production. Nevertheless, as Mann and Repp (1980) and others (Kunisaki & Fujisaki, 1977; Whalen, 1981) have shown, the vowel also exerts an influence on fricative perception: When the fricative noise is ambiguous between /s/ and /ʃ/, listeners report more instances of /s/ when the following vowel is rounded (/u/) than when it is not (/a/), resulting in a quite substantial boundary shift on an /s/-/ʃ/ fricative noise continuum.

A number of other effects of this kind have been found in recent research. For example, a preceding fricative noise (/s/ versus /ʃ/) affects the perception of a following stop consonant (/t/ versus /k/): The /t/-/k/ boundary shifts in favor of /k/ when the precursor is /s/ (Mann & Repp, 1981). The effect is independent of coarticulatory cues to stop place of articulation in

the fricative noise, and it occurs also when the fricative appears to belong to a preceding syllable (Repp & Mann, 1981). Yet another effect operating across a syllable boundary has been obtained by Mann (1980): The boundary on a /da/-/ga/ continuum shifts in favor of /g/ when the preceding syllable is /al/ rather than /ar/.

How are such segmental context effects to be accounted for? Psychoacoustic interactions between adjacent signal portions, while not impossible, become rather implausible. For example, there is little reason to expect that a fricative noise would "sound" different before different vowels. Indeed, when listeners are required to judge the "pitch" of the noise rather than the phonetic category of the fricative, effects of the following vowel disappear (Repp, 1981). The most plausible hypothesis is that segmental context effects represent a perceptual compensation for coarticulatory interactions in speech production. It is well known, for example, that anticipatory lip rounding for rounded vowels affects the noise spectrum of preceding fricatives (Fujisaki & Kunisaki, 1978; Mann & Repp, 1980), and there are indications that the formant transitions of stop consonants shift with the place of articulation of preceding fricatives (Repp & Mann, 1982) and liquids (Mann, 1980). The ability of listeners to compensate for these coarticulatory effects implies an internal representation of these dependencies, which may be conceptualized in dynamic or static terms.

Segmental context effects have been demonstrated even among nonadjacent segments. Thus, shifts in the place of articulation boundaries for initial stop consonants have been found to occur as a function of the place of articulation of the final stop consonant in the same syllable (Alfonso, 1981). Perceptual interdependencies between two vowels separated by a consonant have also been reported (Kanamori, Kasuya, Arai, & Kido, 1971). These effects may reflect perceptual compensation for coarticulatory dependencies operating over wider time spans (cf. Martin & Bunnell, 1981, 1982; Öhman, 1966).

Speaking Rate Effects

The perception of phonetic distinctions that rest on temporal cues may be affected by the temporal structure of surrounding signal portions. Since these effects have been thoroughly reviewed by Miller (1981), we can be brief here.

It is useful to distinguish between experimental manipulations of the duration of selected (steady-state) acoustic segments and of time-varying spectral changes connected with actual (or simulated) changes in articulatory rate. Both temporal and spectro-temporal manipulations have been shown to affect the perception of certain temporal cues, but it is not clear whether their effects take place at the same level.

Some experiments on effects of "speaking rate" concern trading relations among cues for the same phonetic segment. When two temporal cues contribute to the same distinction, a change in one will necessarily require a compensatory change in the other to maintain perceptual constancy. An example of such a trading relation is that between (preceding) silence duration and fricative noise duration as joint cues to the fricative-affricate distinction (Repp et al., 1978). Affricate percepts are favored by both long silences and short noises, so an increase in silence duration can be compensated for, within limits, by an increase in noise duration. But when this trading relation was ex-

amined in the context of a true rate manipulation--the critical cues were embedded in sentence frames produced at a fast or at a slow rate--relatively more silence was needed in the fast sentence frame to maintain the same level of affricate responses. One possible interpretation of this reliable effect (cf. Dorman, Raphael, & Liberman, 1979) is that, in the rapidly articulated context, the (constant) fricative noise sounded relatively longer and hence more fricative-like, so that a longer silence was required to restore the same level of affricate responses. This assumes that the perception of the silence cue was less affected by the rate manipulation. Why this should be so is not clear at present. We should also remark that the speaking rate effect was probably mediated primarily by the immediately adjacent signal portions--the durations of the vocalic segments preceding the silence and following the fricative noise. If so, the speaking rate effects observed may have been a special instance of a segmental context effect or even a trading relation.

A good example of another "speaking rate effect" that could be put, as well, in the preceding section on segmental context effects is the influence of the duration of a following vowel on the perception of the /b/-/w/ distinction (Miller & Liberman, 1979): The longer the vowel, the longer the formant transition duration at the /b/-/w/ boundary. This finding was interpreted as a speaking rate effect, and it is indeed consistent with observed changes in /w/ transition duration at different speeds of articulation (Miller & Baer, 1983). However, the effect has also been obtained with infants (Eimas & Miller, 1980) and with nonspeech stimuli (Pisoni, Carrell, & Gans, 1983), which suggests a possible psychoacoustic origin--i.e., a temporal normalization early in the perceptual process. It is indeed questionable whether changes in the duration of a (steady-state) synthetic vowel are sufficient to convey anything like "speaking rate." Within the context of cue trading relations, both Fitch (1981) and Soli (1982) have been able to separate perceptual effects of vowel duration from effects due to vowel "structure," i.e., more complex spectral changes taking place over time. It is the latter that are more properly viewed as the carriers of information about rate of articulation.

The examples given above illustrate that true "speaking rate effects" are not easy to distinguish from simpler temporal trading relations and local context effects. Moreover, if speaking rate is varied, those changes that occur closest to the target segment will affect its perception most (Summerfield, 1981). In addition, Miller, Aibel, and Green (1984) have recently demonstrated that listeners' overt judgments of speaking rate do not predict the perceptual effects of rate manipulations. On the other hand, considering the extensive speech knowledge that listeners must possess, it seems reasonable to assume that they also have intrinsic knowledge of the acoustic changes that accompany changes in speaking rate and that they "know" how to apply this knowledge in perception. An example of this was also provided by Miller and Liberman (1979) in their study of the /b/-/w/ distinction. When the following vowel was extended by a nonstationary portion containing transitions appropriate for a syllable-final /d/, the effect on the /b/-/w/ boundary was equivalent to that of shortening the steady-state vowel. This paradoxical finding presumably reflects an increase in the perceived rate of articulation due to the additional phonetic segment in the syllable.

Speaker Normalization Effects

Phonetic boundaries along a spectral cue dimension may shift in accordance with the size of the vocal tract that is perceived to be the source of the utterance--that is the hypothesis, at least. As with speaking rate effects, genuine speaker normalization effects are not easy to distinguish from local context effects and spectral trading relations. Moreover, a demonstration of true speaker normalization requires that the test utterance be perceived as coming from a single source (speaker), which is possible only with target segments that are relatively ambiguous as to their source. For these reasons, there are few convincing demonstrations of speaker normalization effects in the literature.

One of the earliest demonstrations was provided by Ladefoged and Broadbent (1957), who showed that synthetic vowel targets were perceived differently in sentence carriers simulating different speakers. This result was replicated with natural speech by Dechovitz (1977). More recently, May (1976) with synthetic speech and Mann and Repp (1980) with natural speech found a shift in the /ʃ/-/s/ boundary when the same fricative noises occurred in the context of vowels produced by different-sized vocal tracts. More experiments along these lines are needed to establish firmly listeners' sensitivity to the static aspects of the perceived speech source.

Semantic and Syntactic Effects

It is a commonplace observation that listeners tend to hear what they expect to hear. Effects of semantic context are ubiquitous in speech perception (Bagley, 1900-1901; Cole & Rudnicky, 1983). However, these effects are generally obtained only when some acoustic information is missing and needs to be "filled in." Apparently, semantic factors can also influence the phonetic boundary on an acoustic continuum characterized by ambiguous (rather than missing) cues.

That such factors can influence the category boundary on a VOT continuum was demonstrated by Ganong (1980). He found that the boundary shifted in favor of word responses when one of the alternatives was a word and the other a nonword, even though the phonetic distinction was in the initial consonant. The pattern of the data suggested that the effect was not merely a response bias; rather, lexical status seemed to influence phonetic categorization directly. But this kind of direct interaction between "top-down" and "bottom-up" processes is a controversial notion (see, e.g., Swinney, 1982), and we do not wish to enter into a discussion of the matter here. Suffice it to point out that phonetic boundaries may be shifted by semantic biases. Such biases can be manipulated not only by changing the lexical status of the target word but also by inducing expectations through preceding sentence context (Garnes & Bond, 1977; Miller, Green, & Schermer, 1982). However, the phonetic boundary shift obtained in that case may be eliminated by selective attention to the target word (Miller et al., 1982), suggesting that semantic processing can be consciously avoided in certain conditions (e.g., when the same materials are repeated over and over). Interestingly, the same study by Miller et al. (1982) also revealed that effects on segmental perception due to the speaking rate of a carrier sentence could not be voluntarily disengaged.

Effects of syntactic boundaries on certain phonetic distinctions have also been reported (Dechovitz, 1979; Price & Levitt, 1983): If the critical cue for the distinction is silence duration (as in the fricative-affricate contrast), more silence is needed if a syntactic boundary is made to coincide with the silence. Although claims have been advanced that this effect can be produced by syntactic structure per se (Dechovitz, 1979), no convincing evidence for such "pure syntax" effects exists so far. Rather, the effects of syntactic boundaries seem to be mediated by the prosodic changes that accompany them. The fricative-affricate boundary may shift depending on whether the preceding word does or does not have clause-final intonation and lengthening (Price & Levitt, 1983; see also Rakerd, Dechovitz, & Verbrugge, 1982). To what extent these effects should be considered merely local context effects or temporal trading relations remains to be seen. In either case, they seem genuinely phonetic rather than psychoacoustic.

Cross-Language Effects

For the purpose of ruling out psychoacoustic factors and establishing that the location of a phonetic boundary is largely determined by factors internal to the listener, cross-language comparisons are most instructive. Languages do differ in their articulatory-acoustic patterns, frequently even for phonetic categories that seem phonemically identical (see Ladefoged, 1983). To the extent that these cross-linguistic differences are captured by a single acoustic speech continuum (and this is not always the case), we should want to know if, in fact, the phonetic boundaries differ for speakers of different languages.

Unfortunately, cross-linguistic studies using the same stimuli and procedures are not very numerous. Among those that do exist, most have dealt with the voicing dimension, as cued by VOT, taking advantage of the fact that languages such as English, French, and Thai make their voicing contrasts in phonetically different ways. While English distinguishes voiced (either prevoiced or voiceless unaspirated) and voiceless aspirated stops, French, Spanish, and Polish contrast prevoiced with voiceless unaspirated stops, and Thai makes both distinctions. The single voicing boundary for English listeners is located in the short-lag values of VOT, between roughly 20 and 40 ms, depending on place of articulation (Lisker & Abramson, 1970). The single boundary for French, Spanish, and Polish listeners, on the other hand, is generally located at shorter lag times, close to zero, and is considerably more variable (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Keating et al., 1981; Williams, 1977). Thai listeners have two boundaries, one in the voicing lead region (where none of the other languages mentioned exhibits any boundary), and the other at voicing lags somewhat longer than in English (Foreit, 1977; Lisker & Abramson, 1970). Thus, native language does seem to influence the location of comparable phonetic boundaries on a VOT continuum, and it certainly determines whether or not a boundary exists at all.

There is ample evidence that discrimination performance is best in the vicinity of a phonetic boundary. Thus, discrimination peaks shift with the phonetic boundaries across languages. Speakers of a language such as Thai have a discrimination peak in the voicing lead region where English listeners' ability to detect differences is extremely poor (Abramson & Lisker, 1970). Another well-known example of such a cross-language difference is provided by the /r/-/l/ contrast, which is easily discriminated by English listeners but nearly indistinguishable for speakers of Japanese, a language that does not

contain these phonetic segments (Miyawaki et al., 1975). For a review of these and related data, see Strange and Jenkins (1978) and Repp (1984).

In view of the flexibility of phonetic boundaries, demonstrations of a coincidence of category boundaries obtained for chinchillas or monkeys with those of English-speaking humans lose some of their impact. To the extent that these animal boundaries are stable at all (see Waters & Wilson, 1976, for a demonstration of large range effects), they may reveal certain psychoacoustic sensitivities that, however, seem to exert only a weak constraint on the possible locations of human boundaries.

It is likely, of course, that the locations of phonetic boundaries in the languages of the world are not totally arbitrary. The structure of the speech production apparatus imposes universal constraints on articulation that may be reflected in a limited number of preferred boundary locations. The hypothesis that human infants may possess some innate sensitivity to these universal potential phonetic boundaries (see Aslin & Pisoni, 1980) has recently gained momentum through the remarkable findings of Werker (1982), who showed that prelinguistic American infants are capable of distinguishing phonetic categories foreign to English, but lose that ability around ten months of age. It has not been conclusively established, however, that these prelinguistic category distinctions are truly phonetic, rather than psychoacoustic, in nature. Exposure to the phonetic distinctions of the native language may merely induce a "speech mode" of listening in the one-year-old infant and thereby lead it to ignore irrelevant acoustic detail. Similarly, several demonstrations of adults' ability to discriminate foreign phonetic categories in certain laboratory situations (MacKain, Best, & Strange, 1981; Pisoni, Aslin, Perey, & Hennessy, 1982) may, at least in part, reflect skills of deploying a nonphonetic mode of processing, and not the acquisition of a new phonetic distinction that can be generalized beyond the laboratory. On the other hand, mastery of a new language does imply the establishment of new phonetic categories, and it is primarily a matter of implementing all the necessary controls to permit the conclusion that this is indeed what has happened in any given laboratory experiment. Rigorous investigations of the process of phonetic learning, which may be a good deal slower than the time span of the typical speech experiment, are just beginning (e.g., Flege & Port, 1981).

Conclusion

Evidence from a variety of experiments on speech perception establishes that phonetic category boundaries are flexible in response to each of two quite different sets of conditions. One set is commonly created by the way utterances are arranged in experiments that require the presentation of sequences of test stimuli. Most of the effects of such conditions are found with nonspeech sounds as well, though, for reasons that are not yet clear, some may be peculiar to speech. The other conditions are the more interesting, at least for our purposes, because they seem to be integral parts of the processes by which utterances are perceived in any test sequence and so, presumably, in the real-life situation. Their effects are of several superficially different kinds, but, common to all, there is a (more or less) apparent correspondence between the shift in the perceived category boundary and the acoustic effects of an articulatory or coarticulatory maneuver. Thus, these boundary shifts imply a link between speech perception and speech production, much as if perception were constrained by tacit "knowledge" of what a vocal

tract does when it makes linguistically significant gestures. Considerations of this kind, roughly similar to those that led originally to the (so-called) "motor theory of speech perception" (Liberman, Delattre, & Cooper, 1952), lead us to suppose that such boundary shifts as these are peculiar to speech.

References

Abramson, A. S., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. Proceedings of the Sixth International Congress of Phonetic Sciences (pp. 569-573). Prague: Academia.

Ainsworth, W. A. (1977). Mechanisms of selective feature adaptation. Perception & Psychophysics, 21, 365-370.

Alfonso, P. (1981). Context effects on the perception of place of articulation. Journal of the Acoustical Society of America, 69 (Supplement No. 1), S93. (Abstract)

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), Child phonology (Vol. 2, pp. 67-96). New York: Academic Press.

Bagley, W. C. (1900-1901). The apperception of the spoken sentence: A study in the psychology of language. American Journal of Psychology, 12, 80-130.

Bailey, P. J., & Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. Journal of Experimental Psychology: Human Perception and Performance, 6, 536-563.

Blumstein, S. E., Stevens, K. N., & Nigro, G. N. (1977). Property detectors for bursts and transitions in speech perception. Journal of the Acoustical Society of America, 61, 1301-1313.

Brady, S. A., & Darwin, C. J. (1978). Range effect in the perception of voicing. Journal of the Acoustical Society of America, 63, 1556-1558.

Braida, L. D., & Durlach, N. I. (1972). Intensity resolution: II. Resolution in one-interval paradigms. Journal of the Acoustical Society of America, 51, 483-502.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. Journal of the Acoustical Society of America, 54, 421-428.

Carden, G., Levitt, A., Jusczyk, P. W., & Walley, A. (1981). Evidence for phonetic processing of cues to place of articulation: Perceived manner affects perceived place. Perception & Psychophysics, 29, 26-36.

Cole, R. A., & Cooper, W. E. (1977). Properties of friction analyzers for /j/. Journal of the Acoustical Society of America, 62, 177-182.

Cole, R. A., & Rudnicky, A. I. (1983). What's new in speech perception? The research and ideas of William Chandler Bagley, 1874-1946. Psychological Review, 90, 94-101.

Cooper, W. E. (1974). Adaptation of phonetic feature analyzers for place of articulation. Journal of the Acoustical Society of America, 56, 617-627.

Crowder, R. G. (1981). The role of auditory memory in speech perception and discrimination. In T. Meyers, J. Laver, & J. Anderson (Eds.), The cognitive representation of speech. Amsterdam: North-Holland Publishing Co.

Crowder, R. G. (1982). Decay of auditory memory in vowel discrimination. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 153-162.

Crowder, R. G., & Repp, B. H. (1984). Single formant contrast in vowel identification. Perception & Psychophysics, 35, 372-378.

Dechovitz, D. (1977). Information conveyed by vowels: A confirmation. Haskins Laboratories Status Report on Speech Research, SR-51/52, 213-219.

Dechovitz, D. (1979). Effects of syntax on the perceptual integration of segmental features. In J. J. Wolf & D. H. Klatt (Eds.), Speech communication papers presented at the 97th Meeting of the Acoustical Society of America (pp. 319-322). New York: Acoustical Society of America.

Diehl, R. L. (1981). Feature detectors for speech: A critical reappraisal. Psychological Bulletin, 89, 1-18.

Diehl, R. L., Elman, J. L., & McCusker, S. B. (1978). Contrast effects on stop consonant identification. Journal of Experimental Psychology: Human Perception and Performance, 4, 599-609.

Donald, L. (1976). The effects of selective adaptation on voicing in Thai and English. Haskins Laboratories Status Report on Speech Research, SR-47, 129-136.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 65, 1518-1532.

Eimas, P. D. (1963). The relation between identification and discrimination along speech and non-speech continua. Language and Speech, 6, 206-217.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. Cognitive Psychology, 4, 99-109.

Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. Science, 209, 1140-1141.

Elman, J. L. (1979). Perceptual origins of the phoneme boundary effect and selective adaptation of speech: A signal detection theory analysis. Journal of the Acoustical Society of America, 65, 190-207.

Fitch, H. L. (1981). Distinguishing temporal information for speaking rate from temporal information for intervocalic stop consonant voicing. Haskins Laboratories Status Report on Speech Research, SR-65, 1-32.

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop consonant manner. Perception & Psychophysics, 27, 343-350.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. Language and Speech, 24, 125-146.

Foreit, K. G. (1977). Linguistic relativism and selective adaptation for speech: A comparative study of English and Thai. Perception & Psychophysics, 21, 347-351.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. Language and Speech, 5, 171-189.

Fujisaki, H., & Kunisaki, O. (1978). Analysis, recognition, and perception of voiceless fricative consonants in Japanese. IEEE Transactions (ASSP), 26, 21-27.

Fujisaki, H., & Shigeno, S. (1979). Context effects in the categorization of speech and non-speech stimuli. In J. J. Wolf & D. H. Klatt (Eds.), Speech communication papers (pp. 5-8). New York: Acoustical Society of America.

Ganong, W. F. III. (1978). The selective adaptation effects of burst-cued stops. Perception & Psychophysics, 24, 71-83.

Ganong, W. F. III. (1980). Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance, 6, 110-125.

Ganong, W. F. III., & Zatorre, R. J. (1980). Measuring phoneme boundaries four ways. Journal of the Acoustical Society of America, 68, 431-439.

Garnes, S., & Bond, Z. S. (1977). The influence of semantics on speech perception. Journal of the Acoustical Society of America, 61 (Supplement No. 1), S65. (Abstract)

Healy, A. F., & Repp, B. H. (1982). Context sensitivity and phonetic mediation in categorical perception. Journal of Experimental Psychology: Human Perception and Performance, 8, 68-80.

Kanamori, Y., Kasuya, H., Arai, S., & Kido, K. (1971). Effect of context on vowel perception. Proceedings of the Seventh International Congress on Acoustics (pp. 37-40). OBudapest.

Keating,, P. A., Mikos, M. J., & Ganong, W. F. III. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. Journal of the Acoustical Society of America, 70, 1261-1271.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. Journal of the Acoustical Society of America, 73, 322-335.

Kunisaki, O., & Fujisaki, H. (1977). On the influence of context upon perception of voiceless fricative consonants. Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo), 11, 85-91.

Ladefoged, P. (1983). Cross-linguistic studies of speech production. In P. F. MacNeilage (Ed.), The production of speech (pp. 177-188). New York: Springer-Verlag.

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. Journal of the Acoustical Society of America, 29, 98-104.

Lahiri, A., Gewirth, L., & Blumstein, S. E. (1984). A reconsideration of acoustic invariance for place of articulation in stop consonants: Evidence from a cross-language study. Journal of the Acoustical Society of America, 76.

Lane, H. (1965). Motor theory of speech perception: A critical review. Psychological Review, 72, 275-309.

Liberman, A. M., Delattre, P. C., & Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. American Journal of Psychology, 65, 497-516.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz, & H.-L. Teuber (Eds.), Handbook of sensory physiology, Vol. VIII: Perception (pp. 143-178). New York: Springer-Verlag.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. Applied Psycholinguistics, 2, 369-390.

Mann, V. A. (1980). Influence of preceding liquid on stop consonant perception. Perception & Psychophysics, 28, 407-412.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. Perception & Psychophysics, 28, 213-228.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. Journal of the Acoustical Society of America, 69, 548-558.

Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. Journal of the Acoustical Society of America, 69, 559-567.

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. Journal of Experimental Psychology: Human Perception and Performance, 8, 473-488.

May, J. (1976). Vocal tract normalization for /s/ and /š/. Haskins Laboratories Status Report on Speech Research, SR-48, 67-73.

Massaro, D. W., & Oden, G. C. (1980). Evaluation and integration of acoustic features in speech perception. Journal of the Acoustical Society of America,

Miller, J. L. (1977a). Nonindependence of feature processing in initial consonants. Journal of Speech and Hearing Research, 20, 519-528.

Miller, J. L. (1977b). Properties of feature detectors for VOT: The voiceless channel of analysis Journal of the Acoustical Society of America, 62, 641-648.

Miller, J. L. (1981). The effect of speaking rate on segmental distinctions: Acoustic variation and perceptual compensation. In P. D. Eimas & J. L. Miller (Eds.), Perspectives on the study of speech. Hillsdale, NJ: Erlbaum.

Miller, J. L., Aibel, I. L., & Green, K. (1984). On the nature of rate-dependent processing during phonetic perception. Perception & Psychophysics, 35, 5-15.

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. Journal of the Acoustical Society of America, 73, 1751-1755.

Miller, J. L., Connine, C. M., Schermer, T. M., & Kluender, K. R. (1983). A possible auditory basis for internal structure of phonetic categories. Journal of the Acoustical Society of America, 73, 2124-2133.

Miller, J. L., Green, K., & Schermer, T. (1982). On the distinction between prosodic and semantic factors in word identification. Journal of the Acoustical Society of America, 71 (Supplement No. 1), S95. (Abstract)

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. Perception & Psychophysics, 25, 457-465.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception & Psychophysics, 18, 331-340.

Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. Psychological Review, 85, 172-191.

Ohman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, 39, 151-168.

Petzold, P. (1981). Distance effects on sequential dependencies in categorical judgment. Journal of Experimental Psychology: Human Perception and Performance, 7, 1371-1385.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. Journal of Experimental Psychology: Human Perception and Performance, 8, 297-314.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. Perception & Psychophysics, 34, 314-322.

Price, P. J., & Levitt, A. G. (1983). Prosody and the /s/-/c/ distinction. Unpublished manuscript.

Rakerd, B., Dechovitz, D. R., & Verbrugge, R. R. (1982). An effect of sentence finality on the phonetic significance of silence. Language and Speech, 25, 267-282.

Repp, B. H. (1980). A range-frequency effect on perception of silence in speech. Haskins Laboratories Status Report on Speech Research, SR-61, 151-166.

Repp, B. H. (1981). Two strategies in fricative discrimination. Perception & Psychophysics, 30, 217-227.

Repp, B. H. (1982). Phonetic trading relations and context effects: New evidence for a phonetic mode of perception. Psychological Bulletin, 92, 81-110.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), Speech and language: Advances in basic research and practice, Vol. 10. New York: Academic Press.

Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. Journal of Experimental Psychology: Human Perception and Performance, 5, 129-145.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative and affricate manner. Journal of Experimental Psychology: Human Perception and Performance, 4, 621-637.

Repp, B. H., & Mann, V. A. (1981). Perceptual assessment of fricative-stop coarticulation. Journal of the Acoustical Society of America, 69, 1154-1163.

Repp, B. H., & Mann, V. A. (1982). Fricative-stop coarticulation: Acoustic and perceptual evidence. Journal of the Acoustical Society of America, 71, 1562-1567.

Rosen, S. M. (1979). Range and frequency effects in consonant categorization. Journal of Phonetics, 7, 393-402.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. Perception & Psychophysics, 30, 309-314.

Samuel, A. G. (1979). Speech is specialized, not special. Unpublished doctoral dissertation, University of California at San Diego.

Samuel, A. G. (1982). Phonetic prototypes. Perception & Psychophysics, 31, 307-314.

Sawusch, J. R. (1977). Peripheral and central processing in speech perception. Journal of the Acoustical Society of America, 62, 738-750.

Sawusch, J. R., & Jusczyk, P. (1981). Adaptation and contrast in the perception of voicing. Journal of Experimental Psychology: Human Perception and Performance, 7, 408-421.

Sawusch, J. R., & Nusbaum, H. C. (1979). Contextual effects in vowel perception I: Anchor-induced contrast effects. Perception & Psychophysics, 25, 292-302.

Sawusch, J. R., Nusbaum, H. C., & Schwab, E. C. (1980). Contextual effects in vowel perception II: Evidence for two processing mechanisms. Perception & Psychophysics, 27, 421-434.

Sawusch, J. R., & Pisoni, D. B. (1974). On the identification of place and voicing features in synthetic stop consonants. Journal of Phonetics, 2, 181-194.

Sawusch, J. R., & Pisoni, D. B. (1976). Response organization and selective adaptation to speech sounds. Perception & Psychophysics, 20, 413-418.

Shigeno, S., & Fujisaki, H. (1980). Context effects in phonetic and non-phonetic vowel judgments. Annual Bulletin of the Research Institute for Logopedics and Phoniatrics (University of Tokyo), 14, 217-224.

Simon, H. J., & Studdert-Kennedy, M. (1978). Selective anchoring and adaptation of phonetic and nonphonetic continua. Journal of the Acoustical Society of America, 64, 1338-1357.

Soli, S. D. (1982). Structure and duration of vowels together specify fricative voicing. Journal of the Acoustical Society of America, 72, 366-378.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America, 64, 1358-1368.

Strange, W., & Jenkins, J. J. (1978). Role of linguistic experience in the perception of speech. In R. D. Walk & H. L. Pick, Jr. (Eds.), Perception and experience (pp. 125-169). New York: Plenum Press.

Studdert-Kennedy, M. (1981). Perceiving phonetic segments. In T. Myers, J. Laver, & J. Anderson (Eds.), The cognitive representation of speech. Amsterdam: North Holland.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. Psychological Review, 77, 234-249.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. Journal of Experimental Psychology: Human Perception and Performance, 7, 1074-1095.

Swinney, D. A. (1982). The structure and time-course of information interaction during speech comprehension, lexical segmentation, access, and interpretation. In J. Mehler, E. C. T. Walker, & M. Garrett (Eds.), Perspectives on mental representation (pp. 151-167). Hillsdale, NJ: Erlbaum.

Waters, R. S., & Wilson, W. A., Jr. (1976). Speech perception by rhesus monkeys: The voicing distinction in synthesized labial and velar stop consonants. Perception & Psychophysics, 19, 285-289.

Werker, J. F. (1982). The development of cross-language speech perception: The effect of age, experience, and context on perceptual organization. Unpublished doctoral dissertation, University of British Columbia, Vancouver, B.C.

Whalen, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. Journal of the Acoustical Society of America, 69, 275-282.

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. Perception & Psychophysics, 21, 289-297.

Footnotes

[1]We are uncertain where to place in the present framework another important class of hypotheses, that of acoustic invariance (Kewley-Port, 1983; Lahiri, Gewirth, & Blumstein, 1984; Stevens & Blumstein, 1978). Sometimes invariant properties are described in terms that suggest a boundary-oriented approach--e.g., when a spectral shape is considered to be either rising or falling. On the other hand, the use of optimal "templates" (Stevens & Blumstein, 1978) suggests a prototype-oriented approach. Since the invariance hypothesis postulates invariant acoustic correlates for linguistic distinctive features, it would seem to permit little flexibility in category boundaries, particularly if the boundaries themselves are taken to be the invariant correlates.

[2]Not all the studies we will cite actually examined boundary shifts. Some studies showed only that the perception of a single ambiguous stimulus could be influenced in one or the other direction. It is safe to infer, however, that, had that stimulus been part of an acoustic continuum, the category boundary on that continuum would have shifted in precisely the same direction.

[3]In the same study, however, sequential contrast was found to be contingent on the perceived phonetic category; i.e., the effect of /spa/ differed from that of /ba/ (Sawusch & Jusczyk, 1981). It is worth noting that, in the selective adaptation paradigm, the adaptors are typically presented at a fast rate that may discourage even covert categorization. Phonetic (judgmental) effects may be contingent upon overt or covert labeling of contextual stimuli.

[4]We call them perceptual functions, rather than perceptual processes, because we believe that these accomplishments of the perceptual system should not be viewed in process terms. In any case, whatever neural or cognitive processes may underly these functions is totally unknown at present.

[5]Although there have been persistent attempts to conceptualize single "invariant" acoustic properties for distinctive features in speech (e.g., Kewley-Port, 1983; Lahiri et al., 1984; Stevens & Blumstein, 1978) these properties never fully capture the phonetically distinctive information. It seems to be a fact to be accepted that what may be a unitary event at the levels of linguistic structure or articulatory planning emerges in a fractionated form at the level of acoustic description.

ON CATEGORIZING APHASIC SPEECH ERRORS*

Betty Tuller+

Abstract. Acoustic studies of voice-onset-time in aphasics' speech suggest that errors of fluent aphasics are misselected phonemic targets, whereas nonfluent aphasics' errors are of articulatory origin. However, we must be cautious when extrapolating a theory from only one measure of articulation. In this experiment, I examined utterances produced by five fluent aphasics, five nonfluent aphasics, and two controls. In the first part of the experiment, I replicated previous voice-onset-time studies. Second, I examined the duration of vowels preceding word-final stop consonants as an index of the consonant's voicing category. The pattern of voice-onset-times produced did not predict the pattern of vowel durations. Thus, voice-onset-time cannot be used to characterize more generally the output of the speaker.

Traditional clinical descriptions of aphasia consider the errors in speech produced by posterior, fluent aphasics to originate at the phonemic or phonological planning levels, whereas phonetic or articulatory errors are thought to be more typical of anterior, nonfluent aphasics (Alajouanine, Ombredane, & Durand, 1939; Luria, 1966; Shankweiler & Harris, 1966). Though it is often difficult to disambiguate so-called planning and execution deficits (or phonemic and phonetic deficits), a fine-grained acoustic analysis has great potential for describing the nature of the underlying speech disorder.

Segmental analyses of aphasic speech have typically proceeded by examining one parameter of the acoustic complex that signals a shift in one phonetic dimension. A commonly used measure is voice-onset-time (VOT), a parameter that distinguishes voiced from voiceless stop consonants in syllable-initial position (e.g., Blumstein, Cooper, Goodglass, Statlender, & Gottlieb, 1980; Blumstein, Cooper, Zurif, & Caramazza, 1977; Freeman, Sands, & Harris, 1978; Hoit-Dalgaard, Murry, & Kopp, 1980; Itoh et al., 1980; but see Shinn & Blumstein, 1983, for an analysis of place of articulation errors). VOT is the

---

*Also Neuropsychologia, in press. A preliminary version of this paper was presented at The Academy of Aphasia, Minneapolis, MN, October 23-25, 1983.
+Also Cornell University Medical College.
Acknowledgment. This work was supported by NINCDS grant NS-17778 to Cornell University Medical College, NINCDS grant NS-13617 to Haskins Laboratories, NIH Postdoctoral Fellowship, and by a grant from the Ariel and Benjamin Lowin Medical Research Foundation. I would like to thank Jason Brown and Ellen Grober for experience testing aphasic patients, Laurel Fais for assistance with acoustic and linguistic analysis, and Hugh Buckingham, Katherine S. Harris, J. A. Scott Kelso, and Leigh Lisker for comments on an earlier version of the manuscript.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

acoustic representation of the time between the burst at release of supraglottal occlusion and the onset of glottal pulsing. For voiced stop consonants in syllable-initial position, glottal pulsing might begin before the release burst, or lag as much as 25 ms after release. In voiceless stop consonants, the onset of glottal pulsing might lag behind supraglottal release by approximately 35-80 ms (Lisker & Abramson, 1964, 1967). In normal English speakers, the actual VOT values vary somewhat as a function of, for example, place of articulation, speaking rate, and phonetic context. Nevertheless, the distribution of VOT values for voiced and voiceless word-initial cognates is bimodal and more or less nonoverlapping, particularly when the words are produced in list form.

In contrast to normal speakers, nonfluent aphasics are reported to produce voiced and voiceless stop consonant cognates having about the same VOT values (Freeman et al., 1978), so that the resulting distribution of VOT is unimodal. These data have been interpreted as indicating that the underlying phonological categories have merged. However, the data are also compatible with the view that these speakers select the correct phonemic targets for production, but the articulation itself is so distorted that the difference between cognates is not maintained (at least on the VOT dimension). Blumstein et al. (1980) attempted to examine this question directly. They operationally defined a production error as an error in selecting the phonemic target when the VOT value of the utterance fell within the range of the opposite voice category, as when a required [b] was produced with a VOT value longer than 35 ms. A production error was considered to be of phonetic origin when its VOT value fell between the normal distributions for the voiced and voiceless categories, as when a required [b] was produced with a VOT value between 15 ms and 35 ms. In accord with previous work, Blumstein et al. found a large overlap of VOT values for voiced and voiceless productions by nonfluent (Broca's) aphasics, suggesting a pervasive deficit in the timing of articulatory movements. They noted, however, that nonfluent aphasics produced some apparent phonemic errors as well, particularly on voiceless stop consonants. In contrast, errors produced by fluent (Wernicke) aphasics tended to fall within the VOT range of the opposite voice category, suggesting that their errors were primarily errors in selecting the appropriate phonemic target, although some apparent phonetic errors were also noted.

This description is intuitively satisfying in that it agrees with subjective clinical impressions. However, as Blumstein et al. recognize, we must be cautious when hypothesizing differences in the mechanisms for production errors from only one measure of articulation. For example, even when restricting discussion to the voicing feature, we find at least sixteen cues that potentially influence perception (Lisker, 1978). If the pattern of errors on the VOT dimension is truly indicative of a more general speech disorder, then some predictions should hold true. Specifically, a speaker producing apparent phonemic errors as reflected in VOT values might be expected to produce a similar distribution of errors when the same phonemic target appears in different positions in a syllable, even though the phonetic realization of the phoneme may be quite different. For example, in English, one strong cue to stop consonant voicing in syllable-final position is the duration of the preceding vowel, which tends to be longer before voiced than before voiceless consonants for both adults (House, 1961; House & Fairbanks, 1953; Klatt, 1973; Peterson & Lehiste, 1960; Raphael, 1975) and children (Raphael, Dorman, & Geffner, 1980). Thus, if the errors are truly of phonemic selection and have no phonetic component, aphasic speakers who produce voicing errors

that fall within the range of VOT values for the opposite voice category in syllable-initial position should show voicing errors in syllable-final position characterized by preceding vowel durations that fall within the range of vowel durations occurring for the opposite voice category (i.e., a bimodal distribution of vowel durations).

The predictions regarding apparent phonetic errors are much less clear. Basically, the number of errors produced should be a function of the difficulty of articulation, which might be affected by a segment's position within a word. Unfortunately, it is as yet impossible to quantify the complexity of articulation involved in producing quite different acoustic results. If, however, the articulations involved in producing changes in VOT and vowel duration are of the same order of difficulty, nonfluent speakers should show the same distribution pattern for voicing errors in syllable-initial and syllable-final position. If voicing production in initial position is more difficult than in final position (as one might perhaps expect from the difficulty aphasics often have initiating speech), we would expect a greater number of phonetic errors in initial position than in final position. Another possibility is that "articulatory complexity" differs across speakers. If this is so, individual speakers might show a coherent pattern of phonetic errors across syllable positions that is not evidenced by the clinical group.

The study reported here is an attempt to determine whether the pattern of production errors indexed by VOT can be used to characterize more generally the output of the aphasic speaker as containing primarily "phonetic" or "phonemic" errors. To this end, the VOT findings of Blumstein et al. (1980) and Itoh et al. (1980) are first replicated. Next, for the same speakers, the duration of the vowel preceding a final stop consonant is examined. Both acoustic dimensions are interpreted with regard to "apparent phonemic" and "apparent phonetic" errors. In this study, errors are operationally defined as "apparent phonemic" errors when categories are misplaced along some acoustic dimension, though contrast is maintained. "Apparent phonetic" errors are operationally defined as those instances of production that fall between categories.

Method

Subjects. The subjects in this study included five fluent (Wernicke) aphasics (referred to hereafter as F1 through F5), five nonfluent (Broca's) aphasics (referred to as NF1 through NF5), and two normal controls. The fluent aphasics were articulatorily agile and used phrases of normal length. However, their speech often made no sense. All of the nonfluent aphasics spoke hesitantly, with long pauses between words, that is, in an "effortful" manner. Three of the nonfluent aphasics would be characterized as agrammatic (NF1, NF2, and NF5) and three were apractic (NF3, NF4, and NF5). The diagnostic category of each patient was determined by performance on the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1972) and other neurological and neuropsychological tests. A list of 35 monosyllabic and polysyllabic words and sentences (selected from a larger list provided by Darley, Aronson, & Brown, 1975) was used to assess the presence of speech apraxia. A speaker was diagnosed as "apractic" if production of the list contained numerous but inconsistent phonetic errors of various types, as well as attempts at self-correction. The errors were judged by a linguist who had no information concerning the individual patients. In all cases, etiology was vascular and involved only the left hemisphere (see Table 1 for additional information).

No tumor or trauma cases were included. All of the subjects were right-handed premorbidly.

Table 1

Descriptive data for aphasic subjects

Speaker Type	Age	Sex	Years of Schooling	Year of Onset	Auditory[a] Comprehension	Hemiplegia
Fluent						
F1	57	F	16	1972	+.7	No
F2	67	F	16	1969	-.3	No
F3	49	M	16	1977	+.06	No
F4	55	M	10	1976	-.12	No
F5	43	M	14	1972	-.6	No
Nonfluent						
NF1[b]	61	F	16	1979	+.2	No
NF2[b]	66	M	12	1980	.0	Yes
NF3[c]	67	M	4	1979	+.7	Yes
NF4[c]	69	M	20	1980	+1.0	Yes
NF5[bc]	52	M	8	1974	+1.0	Yes

[a]Mean of the four auditory comprehension subtests of the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1972); [b]Agrammatism; [c]Speech apraxia

Stimuli. The stimuli were thirty prepausal stressed consonant-vowel-consonant words whose vowel was always [æ]; however, slight vowel quality changes across words did occur for some speakers. The test words included minimal pairs differing on the voicing of either the initial or final consonant (e.g., bat vs. pat and bat vs. bad). Each word (preceded by the word "THE") was printed in large capital letters on an index card and presented to the subject in random order.

Procedure. Subjects were tested individually in a sound-insulated room. On presentation of the stimulus card, subjects were required to read the phrase aloud at least twice. If the subject was unable to read the card easily, the experimenter would pronounce the phrase for the subject to repeat. The randomized list of phrases was presented a minimum of eight times so that each subject attempted to produce at least sixteen tokens of each stimulus word. Subject responses were recorded onto a high-quality tape recorder for later analysis.

Data analysis. Broad phonemic transcriptions of all utterances were made by a trained linguist. Target segments transcribed with a different manner (e.g., [m] instead of [b]) or place of articulation (e.g., [d] instead of [b]) are excluded from further report. Substitutions of, for example, [b^{h}] for [b] were included in the analyses. VOT and vowel duration of the remaining

utterances were measured using an interactive computer program that displays the acoustic waveform. VOT was defined as the time from the energy burst representing initial stop consonant release to the onset of acoustic periodicity representing vocal fold vibration. Vowel duration was defined as the interval from the onset of acoustic periodicity (excluding any initial aspiration) to the first acoustic evidence of closure for the final stop consonant (the time when the high frequency components of the periodic wave disappear). Spectrograms were also used when VOT or vowel duration could not be measured from the acoustic waveform.

Results

Voice Onset Time

The frequency distribution of the VOT values was plotted individually for each subject. Figure 1 shows examples of the distribution of VOT values for a normal control, a fluent aphasic (subject F3), and two nonfluent aphasics (subjects NF2 and NF4). Data from these particular aphasics are shown because F3 and NF2 produced the expected patterns of VOT distribution but NF4 did not. The distributions were analyzed in two ways. First, apparent phonetic and apparent phonemic errors were catalogued using the procedure described by Blumstein et al. (1980). Briefly, if at least two instances of crossover of VOT values between the voiced and voiceless distributions occurred, then all VOT values within this middle range were counted as apparent phonetic errors. The boundaries for this middle range were taken from earlier studies of VOT values in normal speakers (Lisker & Abramson, 1964, 1967) and were +15 to +35 ms VOT for bilabial stops, +20 to +40 ms for alveolar stops, and +25 to +45 ms for velar stops. For a production to be counted as an apparent mistargeting error, its VOT value had to fall in the range appropriate for its voicing cognate.

The results of this analysis are shown in Table 2 and are in fairly good agreement with other reports (Blumstein et al., 1980; Freeman et al., 1978; Hoit-Dalgaard et al., 1983; Itoh et al., 1980). A two-way ANOVA resulted in no significant main effects of group, $F(1,8) = 0.31$, $p > .1$, or error type, $F(1,8) = 0.05$, $p > .1$, but a significant groups by error type interaction, $F(1,8) = 5.69$, $p < .05$. As can be seen from the totals column in Table 2, this interaction occurred because the nonfluent aphasics as a group produced more apparent phonetic than phonemic errors, whereas the fluent aphasics as a group produced more apparent phonemic than phonetic errors. The columns representing the different target sounds indicate that this differential pattern of errors occurred for nonfluent aphasics on all of the six target sounds but on only four of the six target sounds for fluent aphasic speakers. Moreover, the tendency for nonfluent speakers to produce more apparent phonemic errors on voiceless than voiced stops was not replicated. The two control subjects produced no errors of any sort.

Table 3 shows the error patterns for individual speakers. Four of the five fluent aphasics showed mostly bimodal distributions of VOT with the majority of errors falling within the range of the other voice category (apparent phonemic errors). For one fluent aphasic (F4), the voiced and voiceless categories were overlapped considerably, with many errors produced in both the apparent phonemic and apparent phonetic ranges. It is not clear from results of the diagnostic battery why this subject differed so markedly from the other fluent aphasics.

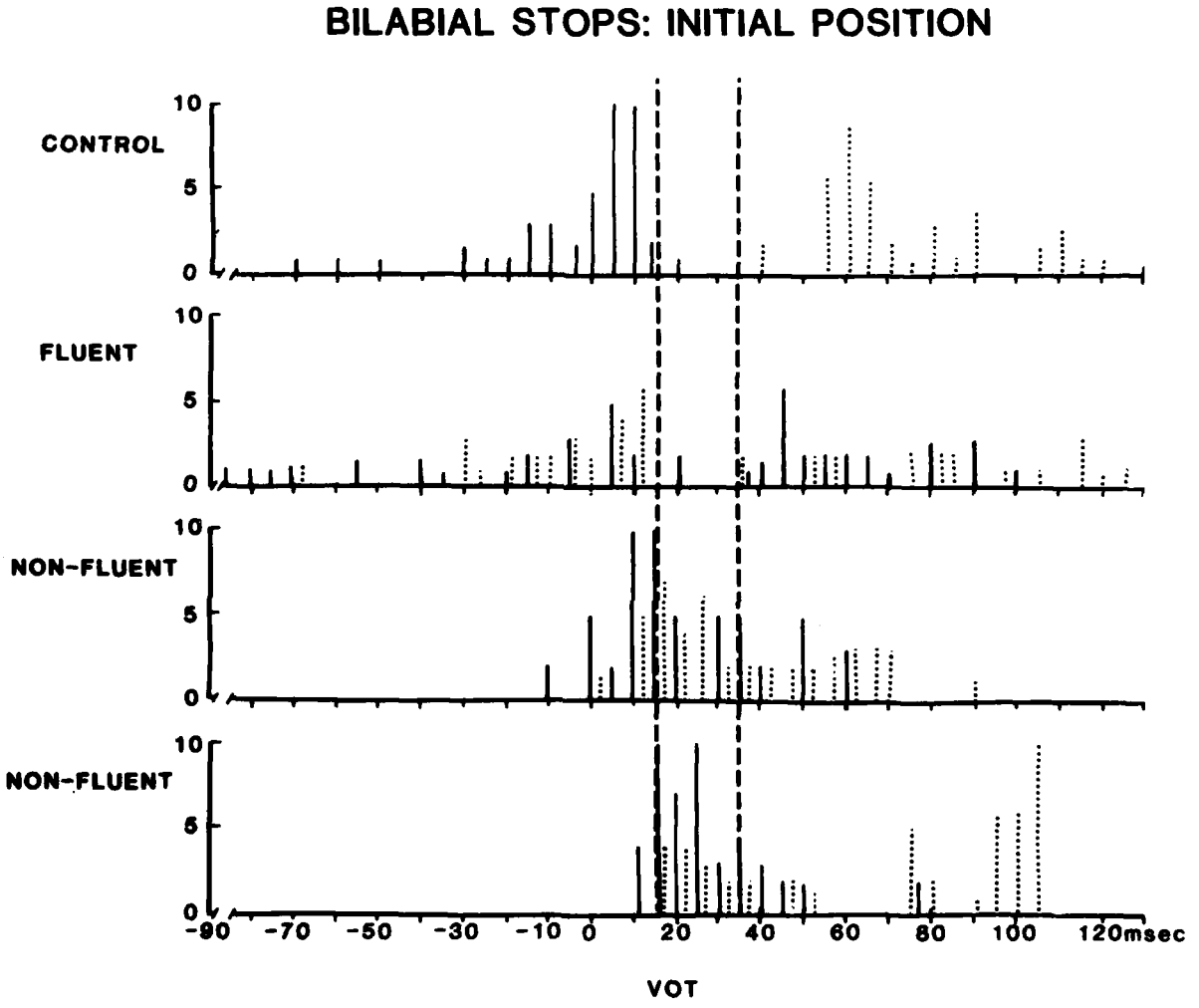


Figure 1. The distribution of VOT values for bilabial stop consonants in word-initial position for a normal control, a fluent aphasic, and two nonfluent aphasics. VOT is plotted on the abscissa of each graph, number of productions on the ordinate. The vertical lines crossing the four graphs at 15 ms and 35 ms represent the upper and lower boundaries of voiced and voiceless bilabial stops, respectively. (-) The required production was [b]; (---) the required production was [p].

Table 2

Apparent "phonemic" and apparent "phonetic" errors expressed as a percent of total productions for each target consonant, across speakers

	Target Consonant						
Error Type	p	b	t	d	k	g	Total
"Phonemic"							
Fluent	22.7	24.6	9.7	8.4	12.5	42.2	20.0
Nonfluent	12.0	15.9	3.2	15.8	2.1	19.6	11.4
"Phonetic"							
Fluent	8.3	8.2	13.7	10.7	3.8	5.5	6.7
Nonfluent	22.9	29.9	13.8	21.8	22.8	23.9	22.3

Table 3

"Phonetic" and "phonemic" errors expressed as a percent of each speaker's total production of each consonant (Criteria: Lisker & Abramson, 1964, 1967)

	Target Consonant						
Error Type	p	b	t	d	k	g	Total
"Phonemic"							
F1	0	0	0	0	0	27.8	4.8
F2	9.1	8.0	21.7	0	0	12.5	8.5
F3	52.8	56.8	5.3	0	20.0	59.1	32.3
F4	20.8	25.6	0	41.2	0	68.9	26.0
F5	30.8	32.4	20.9	0	41.4	42.6	28.0
NF1	3.0	14.3	2.8	23.3	0	38.2	13.4
NF2	13.3	27.8	5.6	40.9	0	47.6	22.5
NF3	33.3	1.8	2.8	6.1	0	5.1	7.9
NF4	0	27.1	4.8	8.0	2.4	5.1	7.9
NF5	10.2	8.5	0	0	8.1	2.0	4.8
"Phonetic"							
F1	0	0	0	0	0	0	0
F2	0	0	0	0	0	0	0
F3	0	0	0	0	8.3	0	1.3
F4	37.5	41.0	18.6	50.0	9.3	26.7	30.5
F5	4.2	0	0	3.4	0	0	1.3
NF1	0	0	0	0	0	0	0
NF2	42.2	37.0	22.2	54.5	33.0	38.1	37.8
NF3	15.4	23.2	5.6	18.2	0	0	10.3
NF4	26.5	63.8	6.0	7.4	20.2	41.3	27.3
NF5	28.2	25.1	33.2	29.1	60.7	40.2	36.1

Although all five nonfluent aphasics produced some VOT values that fell well within the range of the target's voice cognate (apparent phonemic errors), four of the five produced proportionally more errors having intermediate VOT values (apparent phonetic errors). In contrast, one nonfluent speaker (NF1) produced no VOT errors that could be characterized as apparently of phonetic origin.

One shortcoming of this analysis is that it does not accurately reflect a situation in which VOTs of the two voice categories are shortened or lengthened relative to normal, whether or not they overlap. For this reason the VOT data were reexamined to determine simply whether the distribution for a given place of articulation was unimodal or bimodal. If the distribution was unimodal, no delineation of apparent phonetic and phonemic errors could be drawn. If the distribution was bimodal, we determined whether an interval of at least 15 ms without a token separated the two concentrations of data. If so, then all tokens that fell in the opposite voice distribution were termed apparent phonemic errors. If the distribution was strongly bimodal but two or three tokens occurred within the interval between modes, 30 ms of overlap midway between the two modes was ignored when counting apparent phonemic errors. The results of this analysis are shown in Table 4.

Notice first that, as a group, the fluent aphasics still seem to produce more apparent phonemic errors than the nonfluent group (15.0% vs. 4.3%). However, this analysis changes one's conclusions concerning the actual number of targeting errors that occur. For example, in the fourth plot in Figure 1 (subject NF4), the VOT values for voiced and voiceless stop consonants are longer than those measured for normal speakers, so that the aphasic category boundaries do not fall at the normal category boundaries. This does not necessarily mean, however, that the categories have merged. Thus, errors in producing word-initial [p] that appeared in our first analysis to be of phonetic origin appear, with this less stringent criterion, as phonemic errors.

In both analyses of VOT, no errors were produced by the control subjects. Interestingly, the one nonfluent speaker who produced only apparent phonemic errors is severely agrammatic, but would not be characterized as having speech apraxia.

Vowel Duration

The duration of the vowel preceding voiced and voiceless final stop consonants was measured to determine whether the resulting pattern of errors is similar to the pattern of VOT errors. Figure 2 shows examples of the distribution of vowel durations measured for the same normal control, fluent aphasic (F3) and one of the nonfluent aphasics (NF2) shown in Figure 1. However with vowel duration, unlike VOT, one does not have a predetermined cut-off value for accepting a token as correct or in error. Rather than arbitrarily defining a range of durations as apparent phonetic errors, it was determined only whether for a given place of articulation, the distribution of vowel durations was unimodal or bimodal. As in the second analysis of VOT, when bimodal distributions were separated by at least 15 ms, apparent phonemic targeting errors were counted. When seemingly bimodal distributions were not separated by at least 15 ms, the 30 ms between the two distributions were ignored. If the VOT results are indicative of a "phonemic" speech disorder, then those aphasics who produced bimodal distributions of VOT values (primari-

Table 4

Percent of intended target and total productions categorized as "phonemic" errors. Criterion: Bimodality of VOT.

	p	b	t	d	k	g	Total
F1	0	0	0	0	0	27.8	4.8
F2	9.1	8.0	21.7	0	0	12.5	8.5
F3	52.8	56.8	5.3	0	25.0	59.5	33.2
F4	*	*	*	*	*	*	*
F5	33.2	32.4	20.9	2.2	41.4	42.2	28.7
							$\bar{x}$ = 15.0
NF1	3.0	6.1	3.3	19.4	5.9	14.7	8.8
NF2	*	*	*	*	*	*	*
NF3	*	*	8.3	6.1	2.8	5.1	4.3
NF4	36.7	6.4	0	0	2.4	5.1	8.6
NF5	*	*	*	*	*	*	*
							$\bar{x}$ = 4.3

(*) Unimodal distribution.

Table 5

Percent of intended target and total productions categorized as "phonemic" errors on the basis of vowel duration.

	p	b	t	d	k	g	Total
F1	8.6	0	8.6	11.1	3.1	8.6	6.7
F2	8.7	0	0	13.8	4.2	4.2	5.2
F3	*	*	6.7	11.1	*	*	3.3
F4	*	*	3.3	9.4	7.1	0	3.6
F5	0	0	*	*	10.2	12.5	5.7
NF1	*	*	*	*	*	*	*
NF2	12.5	12.8	6.7	7.7	8.7	15.0	10.6
NF3	*	*	*	*	*	*	*
NF4	0	47.4	0	9.4	2.8	26.3	14.5
NF5	0	0	0	12.4	21.3	13.5	7.9

(*) Unimodal distribution.

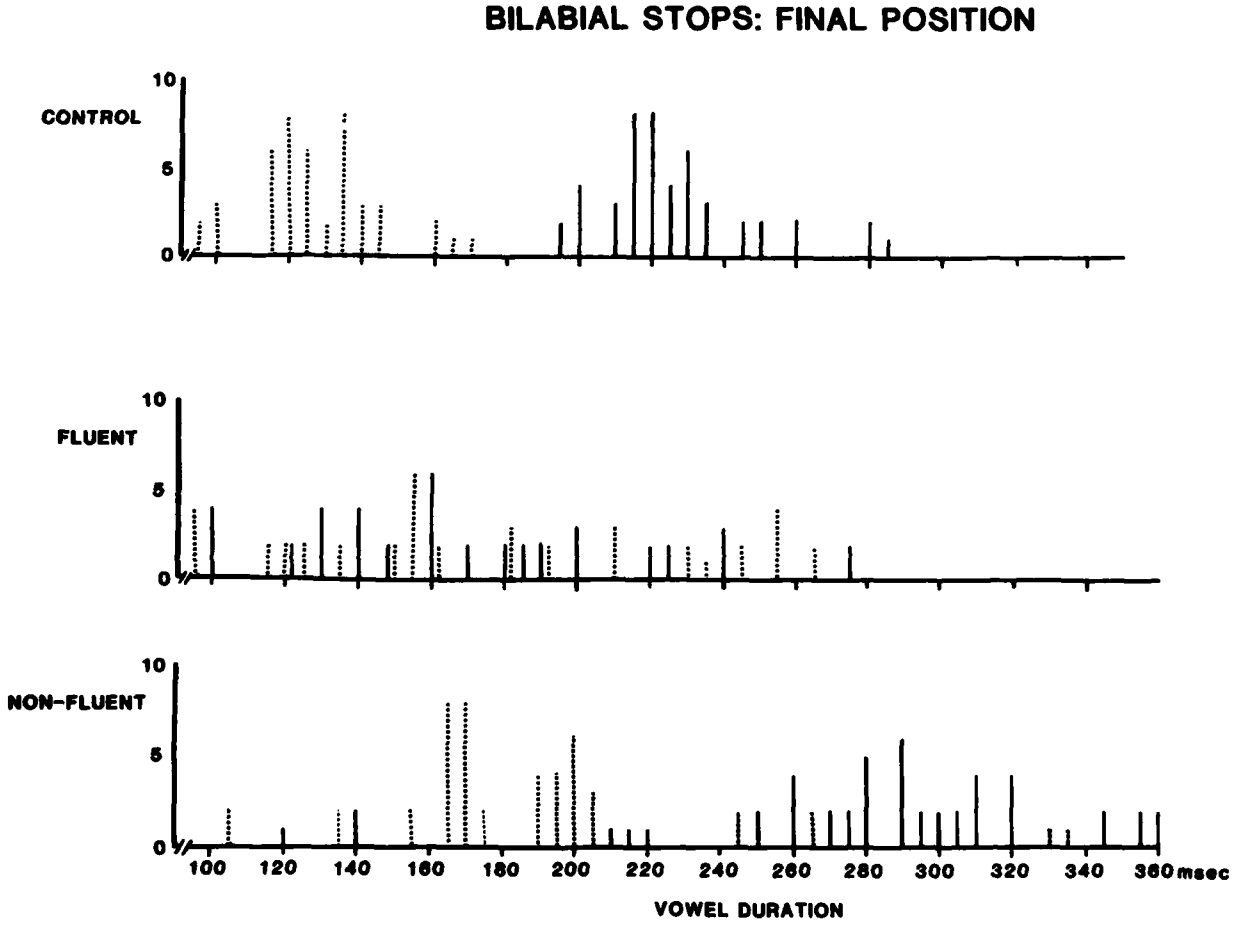


Figure 2. The distribution of vowel duration measures for bilabial stop consonants in word-final position for a normal control, a fluent aphasic, and a nonfluent aphasic. Vowel duration is plotted on the abcissa, number of productions on the ordinate. (-) The required production was [b]; (---) the required production was [p].

ly apparent phonemic errors) should show bimodal distributions of vowel durations.

Table 5 shows the results of this analysis. Notice first that, as a group, the fluent aphasics produced more bimodal distributions of vowel duration (irrespective of number of errors) than did the nonfluent group, although many individual differences within groups are apparent. It is also obvious from comparison of Tables 4 and 5 that the distribution of apparent phonemic errors produced by both fluent and nonfluent speakers is not equivalent for word-initial and word-final positions. Subject F1 produced bimodal distributions of both VOT and vowel duration. Although she produced no apparent phonemic errors on [t] and [d] in word-initial position, in word-final position, 8.6% of her productions in which [t] was the required target, and 11% of productions in which [d] was the required target were apparent phonemic errors. F2 also produced bimodal distributions of both VOT and vowel duration, with the t/d distinction producing the most apparent phonemic errors on both measures. However, in word-initial position the voiced alveolar was substituted for the voiceless, whereas in word-final position the voiceless alveolar substituted for the voiced. Interestingly, this reversal is the consequence of an inappropriate shortening of both VOT and vowel duration. F3 produced bimodal VOT distributions for all three places of articulation, but a bimodal distribution of vowel duration only for the alveolar stops (the category with fewest errors on VOT). F4 produced only unimodal distributions of VOT but bimodal distributions of vowel duration for the velar and alveolar stops. Moreover, the errors in word-initial position greatly outnumbered errors in word-final position. F5 produced many apparent phonemic errors at all places of articulation, as indexed by VOT. In contrast, vowel duration measures indicated apparent phonemic errors only for the velar stops.

With regard to the nonfluent aphasics, one agrammatic speaker (NF1) produced bimodal distributions of VOT values for all places of articulation, but she produced only unimodal distributions of vowel duration. The two other agrammatic speakers (NF2 and NF5) showed the opposite pattern, with unimodal distributions of VOT and bimodal distributions of vowel duration. NF3 produced a unimodal distribution of vowel duration for all places of articulation, but a unimodal distribution of VOT only for bilabial stops. NF4 produced bimodal distributions of VOT and vowel duration values for bilabial, alveolar, and velar stops. However, errors in word-final position occurred predominantly on voiced consonants, a pattern not reflected in word-initial errors. Again, for many of these errors the measured acoustic duration was inappropriately short. As expected, the two normal speakers produced error-free bimodal distributions of vowel duration in these word lists.

In summary, those patients (fluent and nonfluent) who produced apparent phonemic errors in word-initial position did not necessarily produce those errors in word-final position. The result sheds doubt on the conclusion that a production whose value on one acoustic dimension is appropriate to its cognate is indicative of a general impairment in phonemic targeting.

The regularity of apparent phonetic errors can also be questioned given the data in Tables 3, 4, and 5. As previously mentioned, it is possible to demarcate only an arbitrary region of vowel durations, within which productions are categorized as apparent phonetic errors. Thus, a unimodal distribution of measured vowel durations was considered to have "many" apparent

phonetic errors, a bimodal distribution to have "none," and a primarily bimodal distribution with scattered intermediate data points to contain "few" apparent phonetic errors. By these rather loose criteria, no consistency was apparent either within or across speakers. Two of the five fluent aphasics (F1 and F2) produced no apparent phonetic errors on word-initial or word-final stop consonants. Speaker F3 produced only a few apparent phonetic errors on voiceless velar stops in initial position but many apparent phonetic errors on final bilabial and velar stops. Speaker F4 produced many apparent phonetic errors on word-initial stop consonants at all three places of articulation, but only on word-final bilabial stops. F5 produced many apparent phonetic errors only on alveolar stops in word-final position. Of the five nonfluent speakers, two (NF1 and NF3) produced more apparent phonetic errors on final than initial stops. This occurred at all places of articulation for NF1, but only for alveolars and velars for NF3. In contrast, NF2 and NF5 produced many apparent phonetic errors on word-initial stops but none on word-final stops.

Discussion

The results of the first part of this study converge with previous reports of voice-onset-time production by aphasic speakers (Blumstein et al., 1980; Freeman et al., 1978; Hoit-Dalgaard et al., 1983; Itoh et al., 1980). Using the VOT boundaries established by Lisker and Abramson (1964, 1967) it was determined that nonfluent aphasics as a group produced more apparent phonetic than apparent phonemic errors, whereas fluent aphasics as a group produced more apparent phonemic than phonetic errors. It does not necessarily follow, however, that those speakers who produce primarily apparent phonetic errors have merged the voicing categories. When VOT values were examined to determine simply whether the resulting distribution was unimodal or bimodal (ignoring the absolute VOT value), four of the five fluent aphasics and three of the five nonfluent aphasics showed evidence of bimodal patterns. Thus it appears that for these speakers separate voicing categories were preserved.

The major result of this study is that each speaker's pattern of errors on word-initial stop consonants (as measured by VOT values) is not a good predictor of the error pattern on word-final stops (as indexed by vowel duration). For each subject, the number of apparent phonemic errors differed radically across positions. In order to attribute the bulk of the errors produced by fluent aphasics to incorrect selection of phonemic targets, one would have to suppose that the selection of phonemic targets is sensitive to the phoneme's position within a word. There are, in fact, theories that consider a word's representation in the mental lexicon to be phonologically ordered in a left-to-right manner (e.g., Cutler & Fay, 1982; Fay & Cutler, 1977). However, this accounts neither for the unimodal distributions of VOT and vowel duration produced by fluent aphasics nor for the lack of consistency across subjects as to whether more apparent phonemic errors were produced on word-initial or word-final stops.

With regard to apparent phonetic errors, I had hoped to find some consistent pattern, at least for the nonfluent speakers, indicating that adequate control of the interval between release of supraglottal occlusion and the onset of glottal pulsing was more difficult than control of the duration of voicing, or vice versa. However, the pattern and number of errors on initial stop consonant production was unrelated to the pattern and number of errors on final stop production. This may be because 1) the apparent phonetic errors are independent of articulatory complexity, 2) these speakers are brain-dam-

aged so that our (admittedly weak) "metric of articulatory complexity" for normal speakers is not appropriate, or 3) articulatory complexity varies among speakers. Furthermore, the nonfluent speakers did not group on the basis of presence or absence of speech apraxia or agrammatism.

In conclusion, it appears that (at least for this small sample of aphasic speakers) the pattern of errors on the voice-onset-time dimension cannot be used to characterize the total output of the speaker. These data also indicate that the traditional alignment of fluent aphasics with phonemic errors and nonfluent aphasics with phonetic errors is inadequate as a description of aphasic speech production. More generally, we should recognize that phonetic and phonemic aspects of speech are not necessarily independent. Clearly much more acoustic and physiological information is needed before we can ascribe the constellation of fluent and nonfluent aphasic errors to primarily phonetic or phonemic origins.

References

Alajouanine, T., Ombredane, A., & Durand, M. (1939). Le syndrome de la desintegration phonetique dans l'aphasie. Paris: Masson.

Blumstein, S. E., Cooper, W. E., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: A voice-onset time analysis. Brain and Language, 9, 153-170.

Blumstein, S. E., Cooper, W. D, Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset-time in aphasia. Neuropsychologia, 15, 371-383.

Cutler, A., & Fay, D. A. (1982). One mental lexicon, phonologically arranged: Comments on Herford's comments. Linguistic Inquiry, 13, 107-113.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1975). Motor speech disorders. Philadelphia: W. B. Saunders.

Fay, D. A., & Cutler, A. (1977). Malaproprisms and the structure of the mental lexicon. Linguistic Inquiry, 8, 505-520.

Freeman, F. J., Sands, E. S., & Harris, K. S. (1978). Temporal coordination of phonation and articulation in a case of verbal apraxia: A voice onset time study. Brain and Language, 6, 106-111.

Goodglass, H., & Kaplan, E. (1972). The assessment of aphasia and related disorders. Philadelphia: Lea and Febiger.

Hoit-Dalgaard, J., Murry, T., & Kopp, H. G. (1983). Voice onset time production and perception in apraxic subjects. Brain and Language, 20, 329-339.

House, A. S. (1961). On vowel duration in English. Journal of the Acoustical Society of America, 33, 1174-1178.

House, A. S., & Fairbanks, G. (1953). Influence of consonant environment upon the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America, 25, 105-121.

Itoh, M., Sasanuma, S., Tatsumi, I. F., Hata, S., Fukusako, Y., & Suzuki, T. (1980). Voice onset time characteristics of apraxia of speech, part II. Research Institute of Logopedics and Phoniatrics Bulletin, 14, 273-284.

Klatt, D. H. (1973). Interaction between two factors that influence vowel duration. Journal of the Acoustical Society of America, 54, 1102-1104.

Lisker, L. (1978). Rapid vs. rabid: A catalogue of acoustic features that may cue the distinction. Haskins Laboratories Status Report on Speech Research, SR-54, 127-132.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20, 384-422.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. Language and Speech, 10, 1-28.

Luria, A. R. (1966). Higher cortial functions in man. New York: Basic Books.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 32, 693-703.

Raphael, L. J. (1975). The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. Journal of Phonetics, 3, 25-33.

Raphael, L. J., Dorman, M. F., & Geffner, D. (1980). Voicing-conditioned durational differences in vowels and consonants in the speech of three- and four-year old children. Journal of Phonetics, 8, 335-341.

Shankweiler, D. P., & Harris, K. S. (1966). An experimental approach to the problem of articulation in aphasia. Cortex, 2, 277-292.

Shinn, P., & Blumstein, S. E. (1983). Phonetic disintegration in aphasia: Acoustic analysis of spectral characteristics for place of articulation. Brain and Language, 20, 90-114.

UNIVERSAL AND LANGUAGE PARTICULAR ASPECTS OF VOWEL-TO-VOWEL COARTICULATION

Sharon Y. Manuel† and Rena A. Krakow†

Abstract. The present study represents a test of our hypothesis that degree of vowel-to-vowel coarticulation is related to the number and distribution of contrastive vowels in a language. Comparison of vowel-to-vowel coarticulation occurring in Swahili, English, and Shona indicates that there are indeed cross-language differences in the magnitude of coarticulation. Swahili and Shona, which have five-vowel systems, exhibit more coarticulation on vowels than English, which has a considerably larger vowel inventory. The relationship between number of vowels and coarticulation suggests that coarticulation is not simply a by-product of the demands of fluent speech on motor planning and execution. Motor systems, while yielding to the demands of fluent speech, appear to be constrained by the necessity of maintaining distinctiveness, which for each language is defined in the phonology.

Recently, we have been working on a model of coarticulation that focuses on constraints on variability in the acoustic space. In this model we consider phonemes to be associated with target areas as opposed to canonical target points. Coarticulation effects are viewed as a by-product of moving from area to area, rather than deviation from canonical points. It follows from this view that the magnitude of coarticulation should depend upon the size of the target areas.

We hypothesize that there are certain universal principles that tend to constrain the size of target areas, and therefore the magnitude of coarticulation. One obvious candidate for restricting the size of target areas is the need to maintain distinctiveness. We would predict then, that in general, languages with fewer vowels can allocate more space to each vowel area than languages with larger vowel inventories. This hypothesis is based on the premise that the division of the vowel space into distinctive areas is entirely determined by the number of vowels in a particular system. However, while the number of vowels is a major factor in predicting vowel distribution, there

---

†Also Yale University.

Acknowledgment. This research was supported by Grant HD-01994 from the National Institute of Child Health and Human Development. Some of the results were reported at the 58th Annual Meeting of the Linguistics Society of America (Minneapolis, Minnesota, 1983) and at the 106th Meeting of the Acoustical Society of America (San Diego, California, 1983). We are grateful to Suzanne Boyce, Cathe Browman, Carol Fowler, Louis Goldstein, Harriet Magen, Ignatius Mattingly, Patrick Nye, David Odden, Bruno Repp and Michael Studdert-Kennedy for numerous discussions on the subject of coarticulation and for helpful comments on earlier versions of this paper. We would also like to thank Nancy O'Brien for editorial assistance. The order of authors' names is arbitrary.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

are arbitrary, language-particular aspects of vowel distribution that cannot be predicted from any universal principles. Therefore, while we expect that, for example, seven-vowel systems allocate a smaller area to each vowel than three-vowel systems do, this expectation must be qualified by the fact that not all seven-vowel languages use the total vowel-space equally or in the same ways. Within a given language, particular vowels may have more area, and consequently more coarticulatory freedom, in some dimensions than in others. For example, in English there is a binary distinction in the front/back dimension, but several contrastive levels of height. We therefore might expect coarticulation to be freer in the front/back dimension than in the up/down dimension. Keeping in mind all of the language-particular constraints that are not predictable by general principles of distribution, we would still expect the number of vowels in a system to have great predictive power for the size of individual vowel areas. We predict that, in general, languages with smaller vowel inventories can allocate more space to each vowel area than languages with large vowel inventories. If this is the case, and if the size of areas itself determines magnitude of coarticulation, then languages with fewer vowels ought to generally exhibit larger vowel-to-vowel coarticulation effects than languages with more vowels. Essentially, this suggests that coarticulation reflects not only universal motor constraints, but language-particular organization as well.

We have begun to test the specific hypothesis that languages with fewer vowels show more vowel-to-vowel coarticulation than languages with larger vowel inventories. In this paper we present preliminary results from studies of three languages: Swahili, Shona, and English.

Swahili is a five-vowel Bantu language spoken in Southeastern Africa, principally in Kenya and Tanzania. Because Swahili has a smaller vowel inventory than English, we expect Swahili vowels to be more affected by coarticulation than those of English. Based on Öhman's (1966) data, vowel-to-vowel influences appear to be restricted to VC and CV transitions in English and Swedish. If coarticulatory effects are less constrained in Swahili, then we might expect to find that vowel-to-vowel influences extend into the steady state portions of the Swahili vowels.

Swahili has a typical five-vowel system, /i,e,a,o,u/. A male Swahili speaker produced five repetitions each of all possible vowel combinations in VpV and VtV disyllables in a carrier phrase "Nili pata VCV jana" (I received VCV yesterday). In Swahili, the penultimate syllable of a word is stressed, and therefore all VCVs in this experiment were stressed on the first vowel.

Formant trajectories for the vowels were obtained by means of LPC analysis. The values of F1 and F2 in the center of the longest stretch of minimally varying F1 and F2 values were recorded. Figure 1 is a plot of all 100 tokens of each vowel in F1/F2 space, showing a large amount of variability for each vowel. In order to determine how much of this variability is attributable to context, we performed separate four-way analyses of variance on F1 and F2. In each analysis there were four "between" factors: target vowels, flanking vowels, consonants, and positions (first versus second vowel).

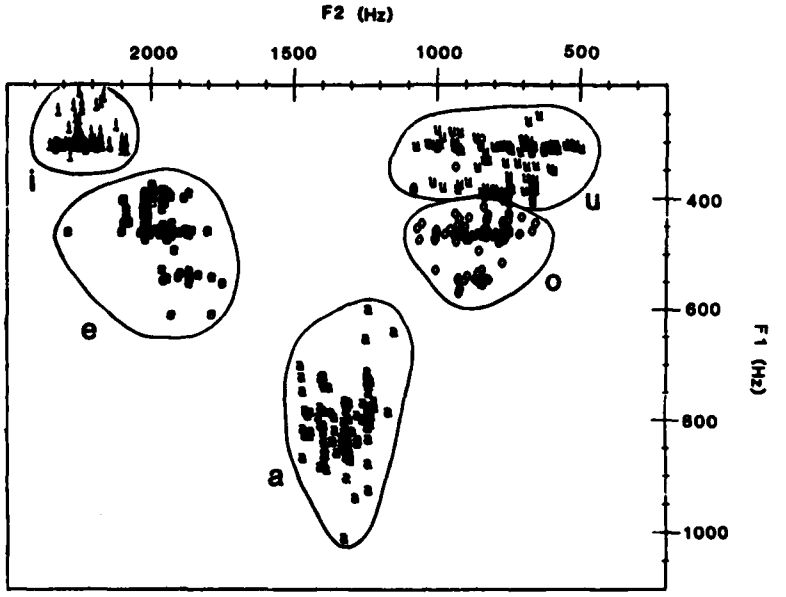


Figure 1. 100 tokens of each of the Swahili vowels in the F1/F2 space.

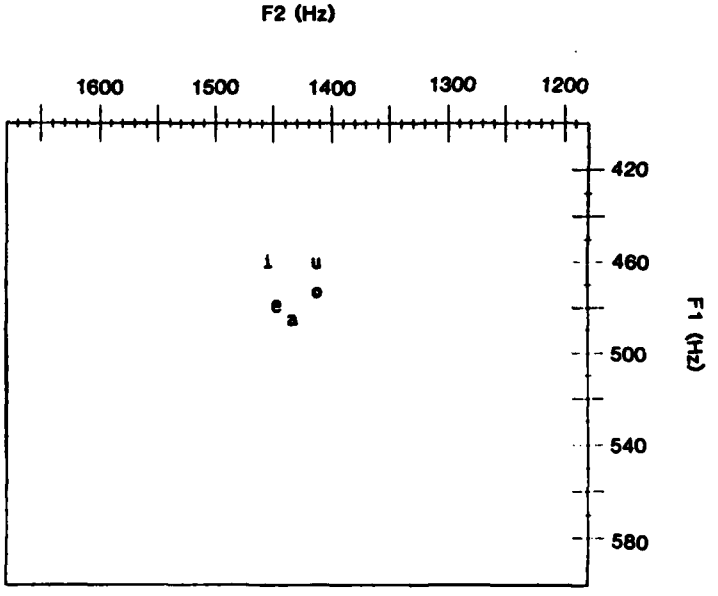


Figure 2. Overall effects of each of the five contextual vowels on the mean F1/F2 values of vowels in Swahili.

Given our hypothesis, we expected that vowels would strongly influence one another across the intervening consonant. Vowels preceded or followed by /i/, for example, should be higher (lower F1 values) and more forward (higher F2 values) than vowels preceded or followed by /a/. In fact, we did find a highly significant and systematic effect of vocalic environment, $F(4,400) = 12.73$, $p < .001$ for F1 and $F(4,400) = 12.97$, $p < .0001$, for F2.

The effects of each of the contextual vowels are shown in Figure 2, in which the means of all vowels for each flanking vowel context are plotted. For example, the symbol "i" is the mean of all vowels when /i/ is flanking, and it shows the effects that /i/ exerts on target vowels. As you can see, F1 and F2 shift in the generally expected directions, so that this figure resembles the Swahili vowel space depicted in Figure 1.

Before we look more closely at how individual target vowels reflect vocalic context, we will examine some other contextual influences on vowels. The effects of intervocalic /p/ versus intervocalic /t/ are shown in Figure 3, in which all target vowels with a particular flanking vowel and a particular intervocalic consonant are plotted. Vowels have a significantly higher F2 in the context of /t/, $F(1,400) = 223$, $p < .0001$. This may reflect forward lingual articulation associated with the /t/ and/or a lowering of F2 in the context of labial /p/. However, if the effect was due to labialization in the /p/ contexts, we would expect F1 as well as F2 to be lowered, but there is no significant effect of consonant on F1, $F(1,400) = 2.27$, $p > .1$. This suggests that the consonant effect is probably due to the lingual movements associated with /t/. Apparent in this figure is the fact that even /t/, which itself involves a tongue gesture, does not block vowel-to-vowel coarticulation.

The amount and type of vowel-to-vowel coarticulation is affected by position. This is shown in Figure 4. The area marked carryover represents the effects of particular first vowels on the mean of all second vowels. The lowercase letters indicate the flanking first vowels and the ways in which they influence the average of all second vowels. The area marked anticipation represents the effects of second vowels on the mean of all first vowels. Anticipatory effects are large, as the figure indicates, and they are statistically significant, in both the F1 and F2 dimensions. On the other hand, carryover coarticulation is significant only for the F2 dimension. (For F1, there was a significant interaction of position by flanking vowel, $F(4,400) = 12.47$, $p < .0001$. Separate ANOVAs for each position revealed a highly significant effect of second vowels on first vowels, $F(4,200) = 27.63$, $p < .0001$. However, vowels in second position were not significantly affected by first vowels, $F(4,200) < 1.0$. For F2, there was no interaction of position with flanking vowel.) Overall, anticipatory coarticulation exceeds carryover coarticulation. This is particularly striking since the first vowel was always the stressed vowel. We will return to this point when we present the data from English.

We will now consider how individual vowels are affected by vocalic context. We will examine the effects of second vowels on first vowels across the medial consonant /p/ in order to simplify the comparison of Swahili with English and Shona, as our data set is limited to VpVs in the latter two languages.

Figure 5 shows the effects of anticipatory coarticulation on each of the five vowels. Within each loop we show the effects of each of the five second vowels on each of the first five vowels. The small letters represent the

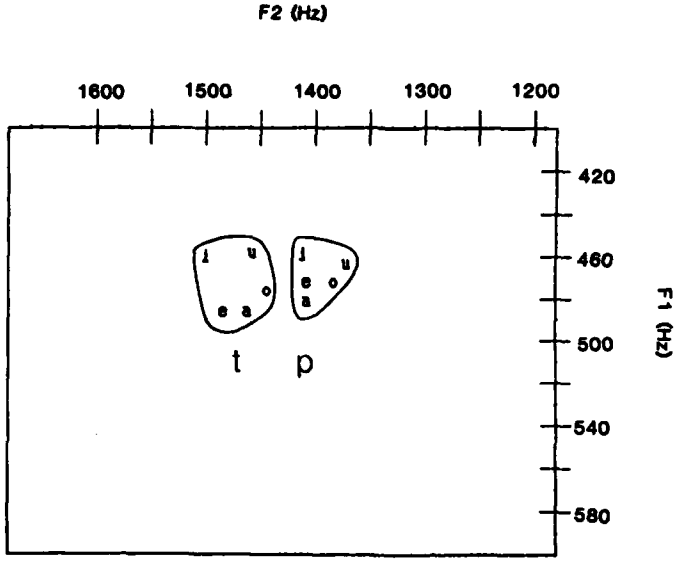


Figure 3. Overall effects of contextual vowels across medial /p/ and medial /t/ on the mean F1/F2 values of vowels in Swahili.

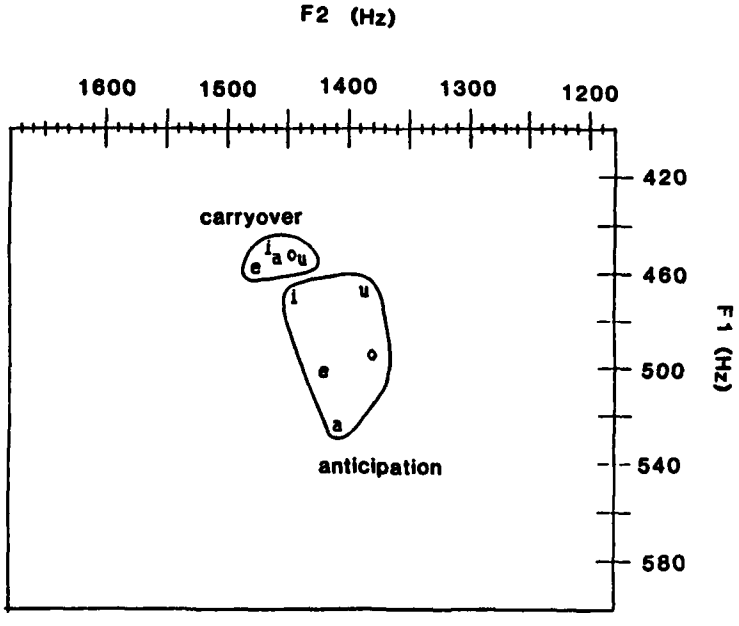


Figure 4. Anticipatory vs. carrryover effects of coarticulation in Swahili.

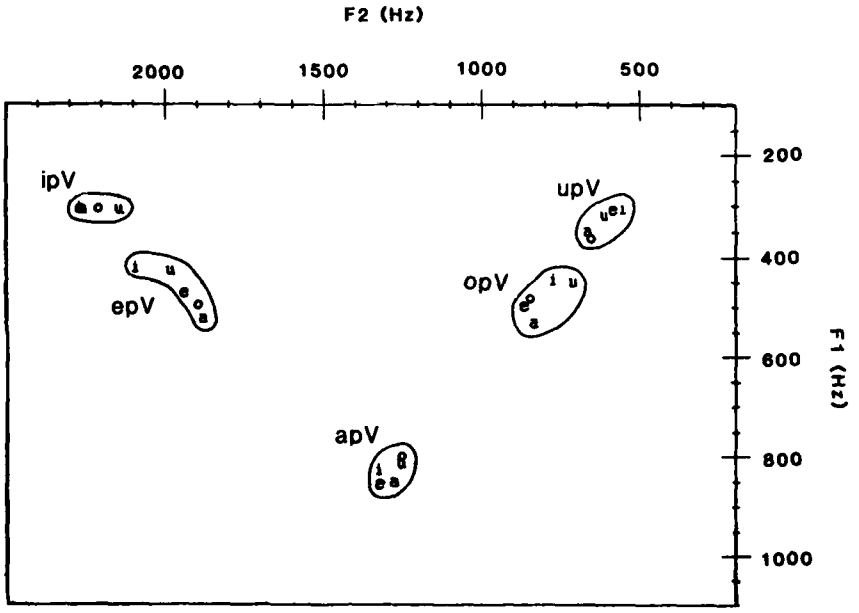


Figure 5. Effects of anticipatory coarticulation on each of the five Swahili vowels.

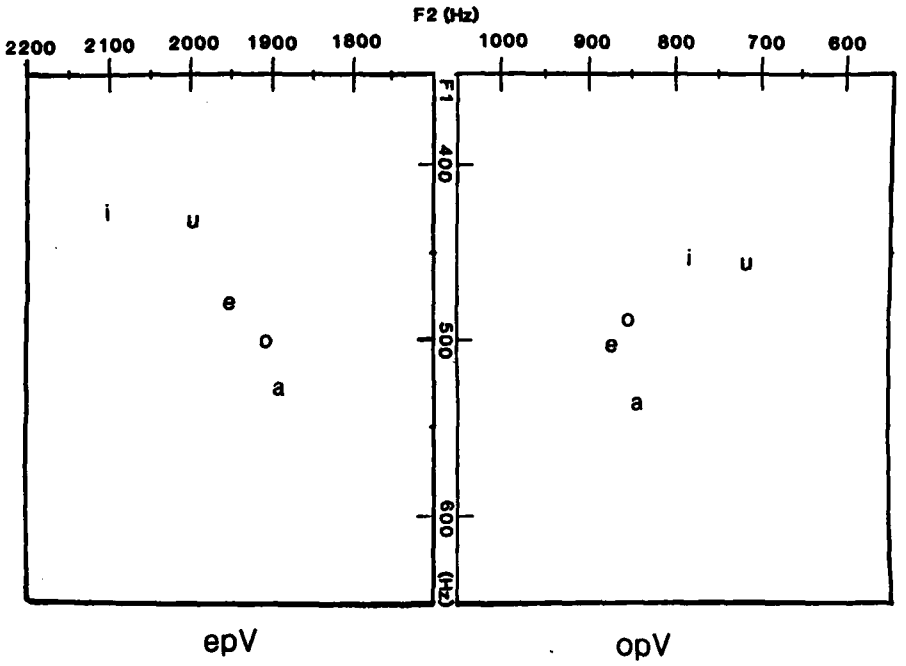


Figure 6. Anticipatory effects of coarticulation in Swahili on the mid vowels, /e/ and /o/.

flanking second vowels. Analysis indicates that the target by flanking vowel interaction is significant for F1, $F(16,400) = 2.25$, $p < .01$ and for F2, $F(16,400) = 4.50$, $p < .0001$. Simple main effects for individual vowels show that /e/, /o/, /a/, and /u/ are significantly affected in both F1 and F2 dimensions, ($p < .05$ in all cases), while /i/ shows significant influence of flanking vowels only in the F2 dimension ($p < .05$). Clearly, anticipatory vowel-to-vowel coarticulation in Swahili is free enough to extend into the steady state portion of each of the vowels.

We now turn to some of the details of vowel-to-vowel effects for other target vowels. The vowel /a/, for example, is affected by the F2 and F1 dimensions of flanking vowels. It appears that the front vowels, /i/ and /e/, pull /a/ forward, with /i/ also raising /a/'s articulation. The back vowels, /o/ and /u/, pull /a/ back and up.

The patterns for the mid vowels are depicted in Figure 6, which is a magnified plot of the effects of flanking vowels on /e/ and /o/. These look almost like mirror images. Clearly the anticipated flanking vowels exert a systematic effect along the height dimension of both mid vowels. The second formant is not affected as we might expect if both backness and rounding are anticipated. Of course, from the acoustics alone it is not generally possible to tease apart the relative contributions of lingual, jaw, and labial gestures.

We have begun to model these coarticulatory effects on an articulatory synthesizer. Preliminary work suggests that much of the acoustic patterning for these vowels can be accounted for by moving the tongue backwards or forwards in anticipation of the upcoming vowel and by moving the jaw in the anticipated direction (which, of course, automatically raises or lowers the tongue along with it). Based on the acoustic data and the articulatory modeling, it appears that these vowels do not reflect the roundedness of the following vowel. It may well be that rounding per se is not contrastive in Swahili. In any case, these data suggest that not all of the configurations of individual articulators are anticipated in the steady state portion of the previous vowel. Nevertheless, the overall vocal tract shape, as reflected in the acoustic data, does show effects of coarticulation.

As predicted, the amount of vowel-to-vowel coarticulation observed for this speaker of Swahili is greater than that reported by Öhman for speakers of English and Swedish. However, Öhman's study of vowel-to-vowel coarticulation was based on spectrographic measures, whereas we used LPC analysis. Although our Swahili data show more extensive coarticulation than Öhman found in English, this could be due to the difference in LPC versus spectrographic analysis procedures. Therefore, we examined English VPV disyllables, using Linear Predictive Coding (LPC) analysis and measuring the most steady state portion of the vowels. As in the case of Swahili, the English VCVs were stressed on the first vowel and embedded in a carrier phrase. The vowels we have analyzed for a single speaker of English are /i, e, a, o/, and the contextual vowels are /i, e, a, o, u/. There were some difficulties in extending our analysis to English since the diphthongal nature of its vowels makes identification of steady state portions somewhat more difficult. Nevertheless, we are confident in the reliability of our measures.

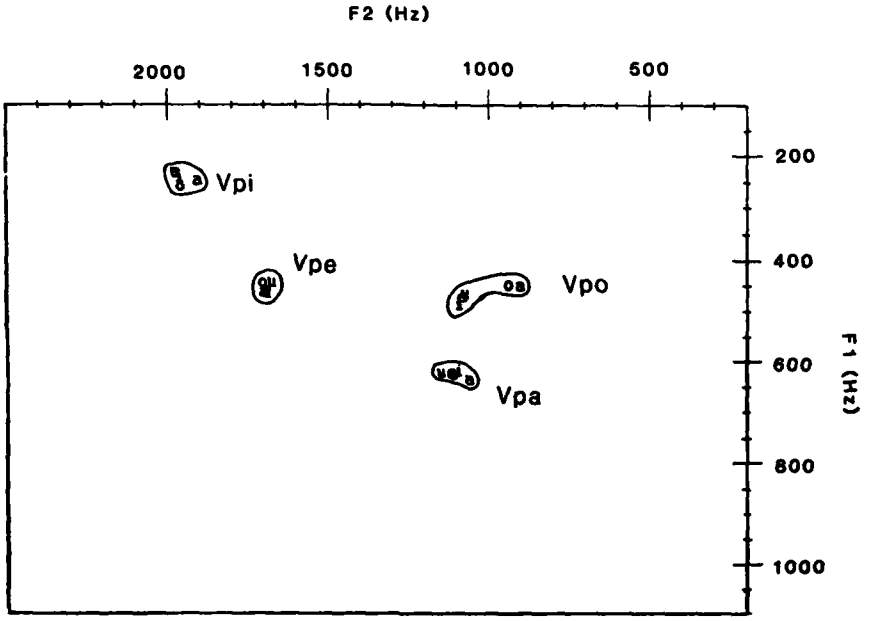


Figure 7. Carryover effects of coarticulation in English.

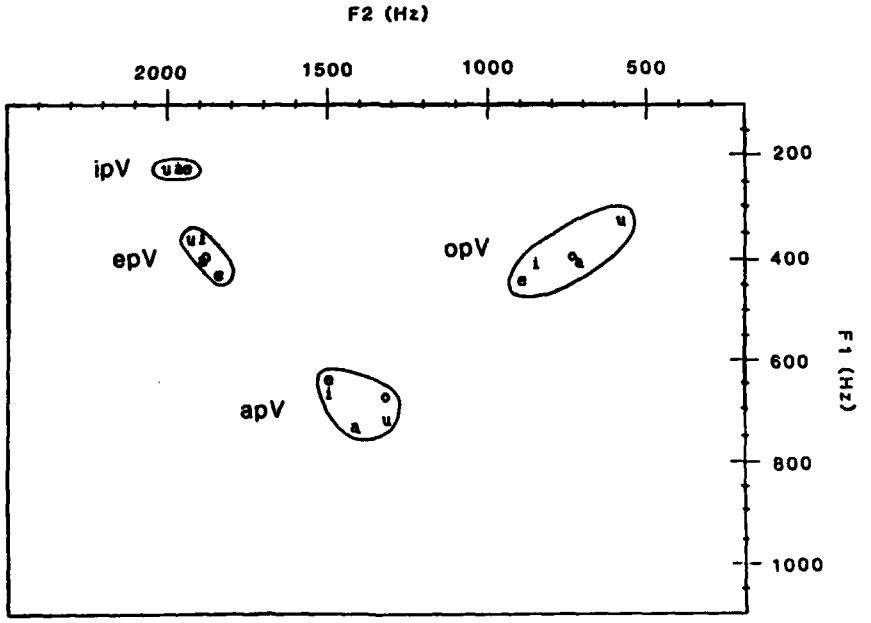


Figure 8. Anticipatory effects of coarticulation in Shona.

Using LPC techniques, we did find some coarticulatory effects in the portions of English vowels that we had identified as most steady state. However, these effects were not as large as those found in Swahili, and were restricted to F2, $F(4,160) = 14.76$, $p < .0001$ (for F1, $F < 1.00$). As discussed earlier, Swahili vowels were significantly affected in both F1 and F2 dimensions. Additionally, anticipatory coarticulation exceeded carryover coarticulation in Swahili. However, in English, carryover effects of coarticulation were significantly greater than anticipatory effects. (There was a significant position by flanking vowel interaction for F2, $F(4,160) = 6.14$, $p < .001$. Separate ANOVAs for each position showed that while for position one, there was a significant effect of the flanking vowel, $F(4,80) = 5.08$, $p < .005$, the effect in second position was much larger, $F(4,80) = 15.54$, $p < .0001$. It may be that directionality of vowel-to-vowel coarticulation is a language-particular phenomenon.

The relative magnitude of coarticulation in Swahili and English can be seen by comparing Figures 5 and 7. In Figure 7 we have plotted the effects of carryover coarticulation in English. As shown in the figure, the effects are small except for the target vowel /o/. The effects are also less regular than in Swahili. In fact, the main effect of flanking vowels on the F2 of target vowels in English may be inflated as a result of the coarticulatory effects exhibited by the target vowel /o/. (There is a significant target by flanking interaction for F2, $F(12,160) = 5.23$, $p < .0001$. Comparing this figure with Figure 5, which shows the anticipatory effects of coarticulation in Swahili, it can be seen that the effects for Swahili are much greater and also seem to be more regular.

We have done the same type of analysis on VPV disyllables in Shona, another five-vowel Bantu language. The magnitude of coarticulatory effects is fairly large in Shona, as shown in Figure 8, in which we have plotted the anticipatory effects of coarticulation. Shona in fact, patterns like Swahili with respect to magnitude of coarticulation. That is, F1 and F2 both show significant effects of coarticulation, $F(4,160) = 3.32$, $p < .05$ for F1, and $F(4,160) = 3.57$, $p < .01$ for F2. Additionally, anticipatory effects exceed carryover effects as we had observed in Swahili.

In summary, comparative analysis of vowel-to-vowel coarticulation in Swahili, Shona, and English supports the hypothesis that, in general, languages with fewer vowels vary more as a function of vocalic context than languages with larger vowel inventories. The number of vowels in a system to a great extent predicts facts about distribution of vowels in the system. However, it is the distribution itself that crucially restricts variation, and there are language-particular determinants of distribution that are not predictable solely by the number of vowels. Thus, for example, in English, which has a relatively large vowel inventory, movement is minimally restricted in the F2 dimension since relatively few vowels occupy the same horizontal plane. We suggest that motor systems, while yielding to the demands of fluent speech, are constrained by the necessity of maintaining distinctiveness. This is a universal principle that results in cross-language variability in coarticulation because distinctiveness is defined for each language in its phonology.

The data presented here provide preliminary support for this hypothesis. We recognize the limitations of generalizing from a single speaker of each of three languages. Clearly, it is necessary to extend this type of analysis to additional speakers and languages. Additionally, it is important to support

the acoustic data with more direct measures of articulatory movement. We have begun to gather additional acoustic data along with articulatory data from several more speakers.

Reference

Ohman, S. E. G. (1966). Coarticulation in VCV utterances: Spectographic measurements. Journal of the Acoustical Society of America, 70, 321-328.

FUNCTIONALLY SPECIFIC ARTICULATORY COOPERATION FOLLOWING JAW PERTURBATIONS DURING SPEECH: EVIDENCE FOR COORDINATIVE STRUCTURES*

J. A. Scott Kelso,† Betty Tuller,†† E. V.-Bateson,††† and Carol A. Fowler††††

Abstract. Speech is surely a complex coordinated activity, but the processes underlying such coordination are not well understood. We show here that articulatory patterns in response to prolonged (1.5 s) and short (50 ms) duration jaw perturbations are not fixed, but are highly specific to the utterance that the speaker produces. In two experiments, an unexpected constant force load (5.88 Newtons) applied during upward jaw motion for final /b/ closure in /bæb/ revealed near-immediate compensation in upper and lower lips, but not the tongue. The same perturbation applied during the utterance /bæz/ evoked rapid and increased tongue muscle activity for /z/ frication, but no active lip compensation. Although jaw perturbation represented a threat to both utterances, no perceptible distortion of speech occurred. That a challenge to one member of a group of potentially independent articulators is met--on the very first perturbation experience--by remotely linked members of the group supports the hypothesis that speech is coordinated through functional synergies (coordinative structures). A third experiment converged on this interpretation by varying the phase of the jaw

---

*Journal of Experimental Psychology: Human Perception and Performance, in press. This paper is a more complete version of one entitled "The functional specificity of articulatory control and coordination," presented at the 104th meeting of the Acoustical Society of America, Orlando, Florida, 8-12th November, 1982 (Journal of the Acoustical Society of America, 1982, 72, S103 [Abstract]). Experiment 3 was presented at the 107th meeting of the Acoustical Society of America, Norfolk, Virginia, 6-10th May, 1984 under the title "Remote and autogenic articulatory adaptation to jaw perturbations during speech: More on functional synergies" (Journal of the Acoustical Society of America, 1984, 75, S23-S24 [Abstract]).

†Also University of Connecticut.
††Also Cornell University Medical College.
†††Also Indiana University.
††††Also Dartmouth College.

Acknowledgment. The work was supported by NINCDS Grant NS-13617, BRS Grant RR-05596, and Contract No. N0014-83-C-0083 from the U.S. Office of Naval Research. Betty Tuller was supported by NINCDS Grant NS-17778. We are very grateful to the following individuals: Dr. Milton Lazanski, former Head of the Department of Orthodontics at Yale University, for his patience in helping us design and produce the first of the jaw prostheses used in this study; Dr. Gerald Alexander for constructing the second; Dr. Kiyoshi Honda for performing the electrode insertions; E. Muller and J. Abbs for technical advice; F. S. Cooper, J. W. Folkins, and four anonymous reviewers for helpful comments on an earlier version.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

perturbation during the production of bilabial consonants. Remote reactions in the upper lip were observed only when the jaw was perturbed during the closing phase of motion, that is, when the reactions were necessary to effect bilabial closure. Thus, coordinative structures are not rigid forms of neuromuscular cooperation; rather, they are flexibly assembled to perform specific functions.

The bewildering complexity of human speech is readily apparent when one attempts to track the spatiotemporal activities of the many anatomical structures involved. One needs little persuasion that talking constitutes an extraordinary feat of motor control, particularly if each degree of freedom were to be individually controlled. A notion that has gained some limited recognition in neuroscience (e.g., Evarts, 1982; Nashner, Woolacott, & Tuma, 1979; Soechting & Lacquaniti, 1981) and behavior (e.g., Bernstein, 1967; Fowler, Rubin, Remez, & Turvey, 1980; Kelso, Southard, & Goodman, 1979; Turvey, 1977) is that the degrees of freedom of any articulator system (however one counts them) are not individually regulated during purposive activity. Rather, in many actions ranging, for example, from locomotion to handwriting, ensembles of muscles and joints exhibit a unitary structuring--a preservation of internal relations among muscles and kinematic components that is stable across scalar changes in such parameters as rate and force (see Grillner, 1982, and Kelso, 1981, for reviews). It appears, then (Bernstein, 1967; Boylls, 1975; Gelfand, Gurfinkel, Tsetlin, & Shik, 1971; Greene, 1972, 1982; Turvey, 1977), that the significant units of control and coordination are functional groupings of muscles and joints (referred to as functional synergies or coordinative structures) that act as a unit to accomplish a task. Therefore, insights into the cooperative behavior among articulators during speech lie in the identification and analysis of coordinative structures.

A window into the behavior of complex systems possessing active, interacting components and large numbers of degrees of freedom can be gained by perturbing them dynamically during an activity and examining how the free variables reconfigure themselves. Thus, a group of potentially independent muscles could be said to comprise a single functional unit if it were shown that a challenge experienced by one (or more) members of the group was responded to by other members of the group at a site remote from the challenge. For the concept of coordinative structure, the response of the articulatory ensemble would not be stereotypic; rather it would be adapted quickly and precisely to accomplish the task. In the case of speech, the components of the neuromuscular apparatus would cooperate in such a way as to preserve the linguistic intent of the speaker.

Although the speech literature contains a number of observations that suggest a coordinative structure mode of articulatory organization, few experiments have employed dynamic perturbation analysis. By and large the "perturbations" introduced to the system have been of a "static" nature. Thus, patterns of cooperation have been observed in various articulators following the fixing of the jaw (as in bite-block experiments, e.g., Fowler & Turvey, 1980; Kelso & Tuller, 1983; Lindblom & Sundberg, 1971), restrictions on lip movements (e.g., Riordan, 1977; Tuller & Fitch, 1980), surgical removal of the alveolar plate or reconstruction of the mandible (e.g., Zimmermann, Kelso, & Lander, 1980), the insertion of palatal prostheses (e.g., Hamlet & Stone, 1978), and so on. Generally, the ability of the speech system to compensate for these disturbances is quite remarkable. However, in many of these studies, various kinds of preadjustments could have occurred before the

test utterances were actually produced. Thus, a more illuminating method may be to perturb the articulators during the speech act and then observe consequent movement patterns, if any, and the speed with which they are achieved.

A pioneering experiment by Folkins and Abbs (1975) did precisely this by occasionally loading the jaw during the closure movement for the initial /p/ in the utterance "a /hæ pæp/ again." Lip closure was attained in all cases, apparently by exaggerated displacements and velocities of the lip closing gestures, particularly by the upper lip.[1] Similarly, Folkins and Zimmermann (1982) used electrical stimulation to produce unexpected depression of the lower lip prior to and during bilabial closure. Compensatory changes in jaw and upper lip movements were observed to effect the bilabial gesture. Although these findings are consistent with the coordinative structure concept, it is not clear from existing data whether, in fact, the patterns of articulator coupling following jaw perturbations are in any sense standardized (as one might predict if they were completely preprogrammed or a result of fixed input-output loops) or whether they are "functional," i.e., directed to the stable production of the intended utterance. If the former, the pattern of response to a given jaw perturbation should be the same regardless of utterance. If the latter, different patterns of articulator cooperation (coordinative structures) should occur, tailored to the particular phonetic requirements.[2]

In the first two experiments reported here, we examined the effects of jaw perturbation on production of two phonetic segments, /b/ and /z/. For /b/, the primary vocal tract constriction is created normally by bilabial closure. For /z/, the main constriction is produced by positioning the tongue in close approximation to the palate or teeth. Note that from a low vowel environment jaw and lips cooperate for production of /b/, whereas jaw and tongue cooperate in the raising gesture for /z/. Thus if the jaw is perturbed during the transition into the final /b/ in /bæb/, then the primary response should occur in the lips, rather than, say, the tongue. In contrast, if the same perturbation is applied during the jaw raising for the final /z/ in /bæz/, the primary response should occur in the tongue, not the lips. Experiment 1 presents an initial exploration of this idea. Experiment 2 provides more detailed electromyographic and kinematic evidence for task-specific articulator cooperation. A third experiment attempts to converge on the interpretation of the first two experiments by examining remote reactions to jaw perturbation as a function of the phase of jaw motion at which loads are applied. For example, upper lip responses should only be observed when the jaw is perturbed during the closing gestures for bilabial consonant production, that is, when the upper lip contributes to vocal tract occlusion.

Experiment 1

Subject, Materials, and Procedures

One adult male (one of the authors) participated in the first two experiments reported here.[3] The speech sample contained two utterance types, "a /bæb/ again" and "a /bæz/ again." In the first part of the experiment, 30 trials of each utterance were performed in a single block. On 20% of the trials (6 randomly selected trials out of 30 for each utterance) a load perturbation was applied to the jaw during the closing gesture for the second consonant, /b/ or /z/. The perturbation was triggered during /bæb/ and /bæz/ when the jaw reached the same predetermined point approximately midway through its upward trajectory. The experiment was performed with a constant force

load of 1.5 s duration. Exactly the same procedure was repeated in the second part of the experiment, but with a 50 ms load. It is important to note that the subject did not know on which trials he would be perturbed. Moreover, until the first perturbed trial, the subject was unaware of the specific locus of the perturbation during the raising trajectory and the magnitude of the applied load.

Apparatus and Data Recording

Figure 1 illustrates the experimental set-up. The subject sat in a dental chair with his head fixed in a specially designed cephalostat (basically a plaster cast mold constructed for the subject's head and a clamp that fitted onto the bridge of the subject's nose--all enclosed in a wooden box, Figures 1A and 1B). A custom-made titanium dental prosthesis fitted onto the subject's lower teeth (Figure 1C). Two small rods of the prosthesis protruded from the sides of the mouth and were coupled by a thin wire to a Brushless DC torque motor that was situated perpendicular to the subject's chin. A load cell placed in series with the coupling wire monitored applied torque. This enabled us to control the torque motor under force feedback and made it possible to couple the motor to the jaw with a very small tracking load of approximately 30 g. Jaw movements were monitored by a rotary voltage displacement transducer placed at the axis of rotation of the sector arm (see Figure 1B). The existence of the tracking force had no perceptible effects on the subject's speech, nor on observed movement and EMG activity. The experiments were completely controlled by a programmable microcomputer that specified on which trials the load was to be added and the magnitude of the load. In each experiment the load was the same (5.88 Newtons). and the rise-time to peak load was small, on the order of 2-3 ms.

Infrared light-emitting diodes were attached at the vermilion border of the subject's upper and lower lips at the midline, and sensed by an optical tracking system (a modified SELSPOT system). The displacements of the articulators and the acoustic speech signal were stored on FM tape for later computer analysis. A set of software routines was used to differentiate the movement signals and display the audio output along with movement information in a time-synchronized format. The acoustic recordings were inspected to determine the first evidence of bilabial closure for the final /b/ in /bæb/ trials (defined here as the point when the high frequency components of the periodic wave disappear) and of frication onset for /z/ in the /bæz/ trials (defined as the onset of high frequency, low amplitude noise).

Results and Discussion

In this experiment, we evaluated the effect of the jaw perturbation on upper and lower lip movement, and whether the effect was context-sensitive. We first established that the 1.5 s load prevented the jaw from reaching its usual position, by measuring jaw height at the earliest acoustic evidence of lip closure or frication. The results are presented in Table 1, which shows the mean articulator positions for the jaw, lower lip plus jaw, and upper lip, obtained from an arbitrary reference point. For both phonetic contexts, the jaw was significantly lower during 1.5 s load trials than for the immediately preceding unloaded trials, $t(10) = 26.99$, $p < .001$ and $t(10) = 3.18$, $p < .05$, for /bæb/ and /bæz/, respectively.

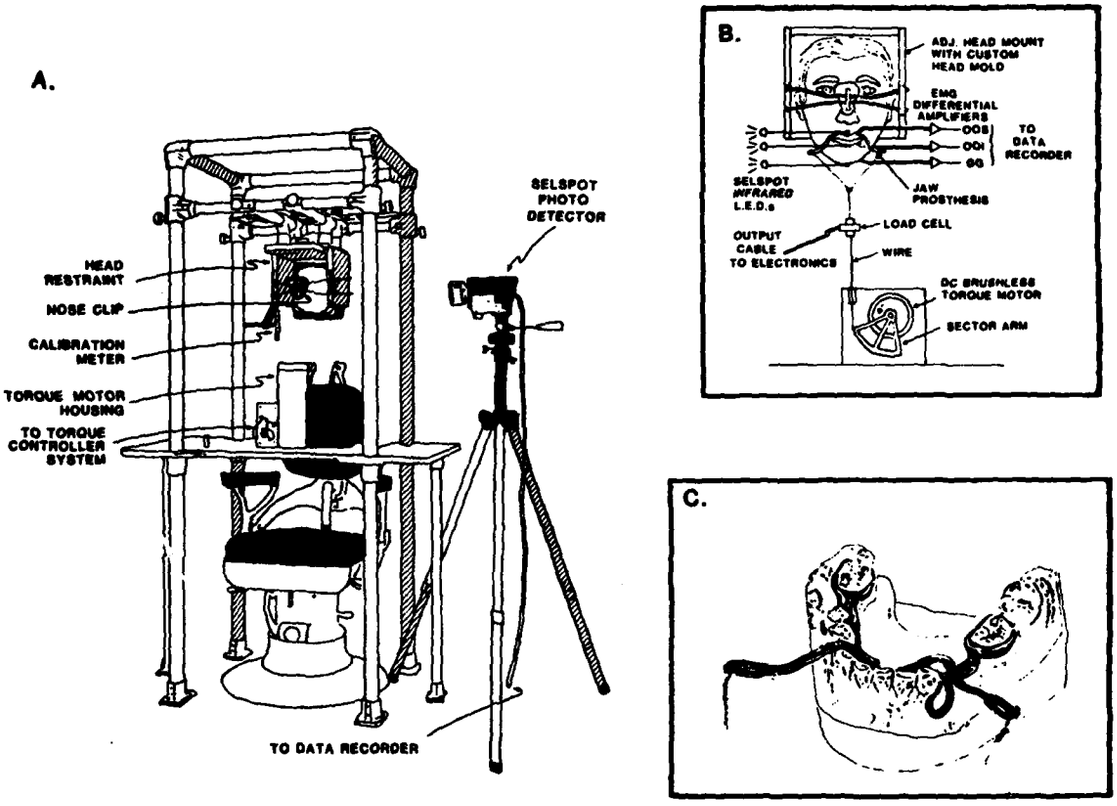


Figure 1. A. The general experimental set-up. B. A schematic of the subject in the head apparatus, showing placement of LEDs for movement tracking and electrodes for monitoring EMG activity. OOS and OOI are orbicularis oris superior and inferior, respectively. GG is the genioglossus, a major tongue muscle (see text for details). C. A specially designed jaw prosthesis. Note gaps for missing teeth that afford a unique capability for setting the prosthesis firmly in the mouth of the subject (see Footnote 3).

Table 1

Mean Articulator Position (and sd) in mm[1]

	At Onset of Closure /bæb/		At Onset of Frication /bæz/	
Load: 1.5 s	Control	Load	Control	Load
Jaw	41.4 (.41)	35.4 (.35)**	42.3 (.01)	34.6 (.01)**
Lower lip	23.3 (.01)	23.0 (.46)	23.3 (.01)	22.8 (.46)*
Upper lip	2.3 (.39)	1.6 (.47)*	6.1 (.24)	5.8 (.41)
50 ms				
Jaw	41.2 (.41)	40.5 (.81)	42.2 (.01)	41.9 (.42)
Lower lip	23.1 (.23)	23.0 (.23)	23.3 (.23)	23.3 (.01)
Upper lip	2.6 (.18)	2.1 (.36)*	5.5 (.44)	5.6 (.39)

*$p < .05$
**$p < .001$
[1]Measured from an arbitrary reference position. The lower the number for a given articulator, the lower is its spatial position.

The coordinative structure concept predicts one consequence of this difference in jaw height, namely, that upper lip displacement downward should increase when producing /b/, but not /z/, when the jaw load is applied. The displacement of the upper lip downward in each trial was measured at the time of acoustic onset of final /b/ closure or final /z/ frication. As predicted, the position of the upper lip at final /b/ closure was lower for the perturbed trials than for the immediately preceding unperturbed trials, $t(10) = 2.64$, $p < .05$. In contrast, there was no difference in upper lip position for /z/ with and without a load, $t(10) = 1.44$, $p > .1$. In addition, the position of the lower lip in space at the point of closure for /b/ was unaffected by the 1.5 s load, indicating a considerable adjustment for the lower jaw position, $t(10) = 1.65$, $p > .1$. Similarly, for /z/ although the lower lip is lower in space, $t(10) = 2.68$, $p < .05$, the difference is small compared to the much lower jaw position. These lower lip reactions will be considered in more detail in the following experiment.

When the applied load was of 50 ms duration, no effect of perturbation was apparent on jaw position by the time closure or frication was achieved, $t(10) = 2.02$ and 1.57 for /bæb/ and /bæz/, respectively, $ps > .05$. Lower lip position also showed no effect of the 50 ms load, $t(10) = 1.05$ for /bæb/ and .42 for /bæz/, $ps > .1$. Although upper lip position for /z/ was similarly unaffected by this short-duration load, $t(10) = 0.26$, $p > .1$, the upper lip in /b/ did increase its downward deflection in loaded trials relative to unloaded trials, $t(10) = 2.96$, $p < .05$. The change in upper lip displacement, but not lower lip, is most probably a function of an increase in compression of the upper lip.

To summarize, these preliminary observations suggest that a disruption in movement of one articulator (the jaw) is responded to by another, remote articulator (the upper lip), when the phonetic context is one for which that reaction is functionally appropriate. However, the experiment has three shortcomings. First, although we have provided evidence of a coordinative structure during /b/ production, we have not provided direct evidence for its presence in /z/ production. Second, in order to understand the articulatory system's response to perturbation, both detailed kinematic and electromyographic information are desirable. Third, and relatedly, in order to evaluate the reliability of the effects described in Experiment 1, a greater number of trials is warranted. For example, in Experiment 1 it may be that the 50 ms load had a slight effect on articulatory movements (as suggested by the increase in upper lip displacement for /b/), but six loaded trials are insufficient to comprise a sensitive enough test. For these reasons, we performed a second experiment, similar in many respects to Experiment 1. In Experiment 2, the total number of trials was increased and, in addition to monitoring jaw and lip movements, electromyographic (EMG) potentials from tongue and lip muscles were obtained. We were especially interested in evaluating tongue muscle activity during /z/ production.

Experiment 2

Subject, Materials, and Procedures

The same subject who participated in Experiment 1 took part in this study. The speech sample contained the same two utterance types as in Experiment 1, "a bæb again" and "a bæz again." In each part of the experiment, 40 trials of each utterance were performed in two 20-trial blocks. At least 5 s separated individual trials. On 25% of the trials (10 randomly selected trials out of 40 for each utterance) a load (5.88 Newtons) was applied to the jaw during the closing gesture for the second consonant, /b/ or /z/. The load was triggered during /bæb/ and /bæz/ when the jaw reached the same predetermined point approximately midway through its upward trajectory. Once again, the subject knew that some of the trials would be perturbed but not which ones. Nor did the subject experience any form of loading (except the tracking load) until the experiment proper. The first part of the experiment was performed with a constant force load of 1.5 s duration, the second part with a 50 ms load. The utterance order was counterbalanced across loading conditions.

Apparatus and Data Recording

The jaw loading device and the methods of tracking movements of the jaw, upper lip, and lower lip were identical to the previous experiment. In addition to these kinematic measures, EMG potentials from a muscle in the upper lip (orbicularis oris superior, OOS) and a muscle in the lower lip (orbicularis oris inferior, OOI) were obtained using paint-on electrodes, while EMG potentials from a tongue muscle (the posterior portion of genioglossus, GG) were obtained using bipolar hooked-wire electrodes inserted by our resident laryngologist, Dr. Kiyoshi Honda. The genioglossus recordings were used as an index of tongue activity during /z/ production. The displacements of the articulators, EMG from tongue and lip muscles, and the acoustic speech signal were stored on FM tape for later computer analysis. Software routines were used to differentiate the movement signals, ensemble average the rectified EMG signals, and display the audio output synchronized with movement and EMG information.

Results and Discussion

First we established once more that the upward jaw trajectory differed in loaded and unloaded trials. The position of the jaw in each trial was measured at the earliest acoustic evidence of final /b/ closure or /z/ frication. Position of the jaw in loaded trials was then compared to normal conditions and was significantly lower for both /bæb/, $t(18) = 10.20$, $p < .001$ and /bæz/, $t(18) = 22.45$, $p < .001$. In Figure 2, a sample of the jaw velocities is shown for the first eight perturbed trials of both /bæb/ and /bæz/ utterances for one subject. The effect of the load perturbation was to alter the direction of jaw movement almost immediately in a very consistent manner. That is, the jaw velocity became sharply negative just after torque onset. Loaded trials show very small trial-to-trial variability in the jaw velocity profiles for both utterances.

The displacements and velocities of the upper lip, the lower lip (with the contribution of jaw subtracted out), and the jaw itself are shown for perturbed and unperturbed ("control") trials in Figures 3 and 4. Each trace represents the average of 10 tokens, with the dotted trace indicating the control utterances and the solid trace the perturbed utterances. The vertical line in each window of the figures marks the onset of torque to the jaw. Even though the torque prevented normal upward jaw motion, lip closure for /b/ and frication for /z/ were attained on all trials. In /bæb/, for example, peak lower and upper lip displacement occurred on the average 5 ms before and 5 ms after acoustic closure,[4] respectively, on control trials, and 11 ms and 7 ms on the average after acoustic closure on perturbed trials. Thus, the timing differences among articulators were small between perturbed and unperturbed utterances, and we were not able to hear any obvious differences in the utterances between the two conditions.

Examination of the kinematics in Figures 3 and 4 and corresponding rectified and averaged EMG in Figure 5 reveals interesting adjustments in response to jaw perturbation. Figure 3A shows that the downward displacement of the upper lip in /bæb/ is greater than its unperturbed control. Measured at the acoustic onset of final /b/ closure, this difference is highly significant, two-tailed $t(18) = 3.19$, $p < .01$. In contrast, for /bæz/ (Figure 3B) the upper lip shows no displacement differences between perturbed and control conditions, $t(18) = .001$, $p > .1$, when measured at the onset of /z/ frication.

One anomalous result is that OOS (Figure 5 top) shows an active increase in EMG activity with an average latency of 20 ms in response to the added load for both /bæb/ and /bæz/ (SD = 18 ms). Thus, even though there are differential movement effects in /bæb/ and /bæz/ as a function of perturbation, the EMG response, at least in terms of its timing, is similar in both utterances. Although puzzling, several, perhaps related, interpretations of this result are possible. One is that although in /bæz/ there was little vertical upper lip displacement, the subject was observed to protrude the lips slightly, a maneuver that could be revealed by measuring horizontal displacement. The present study, however, does not allow us to evaluate this possibility. Relatedly, there are some suggestive hints in the data shown in Figures 4 and 5 that the jaw and upper lip may be functionally coupled in /bæz/ as well as /bæb/. The increase in EMG that is time-locked to jaw perturbation, combined with a small increase in upper lip downward velocity (Figure 4B), render this interpretation viable. Alternatively, the EMG response to perturbation in both /bæb/ and /bæz/ may only reflect a general stiffening

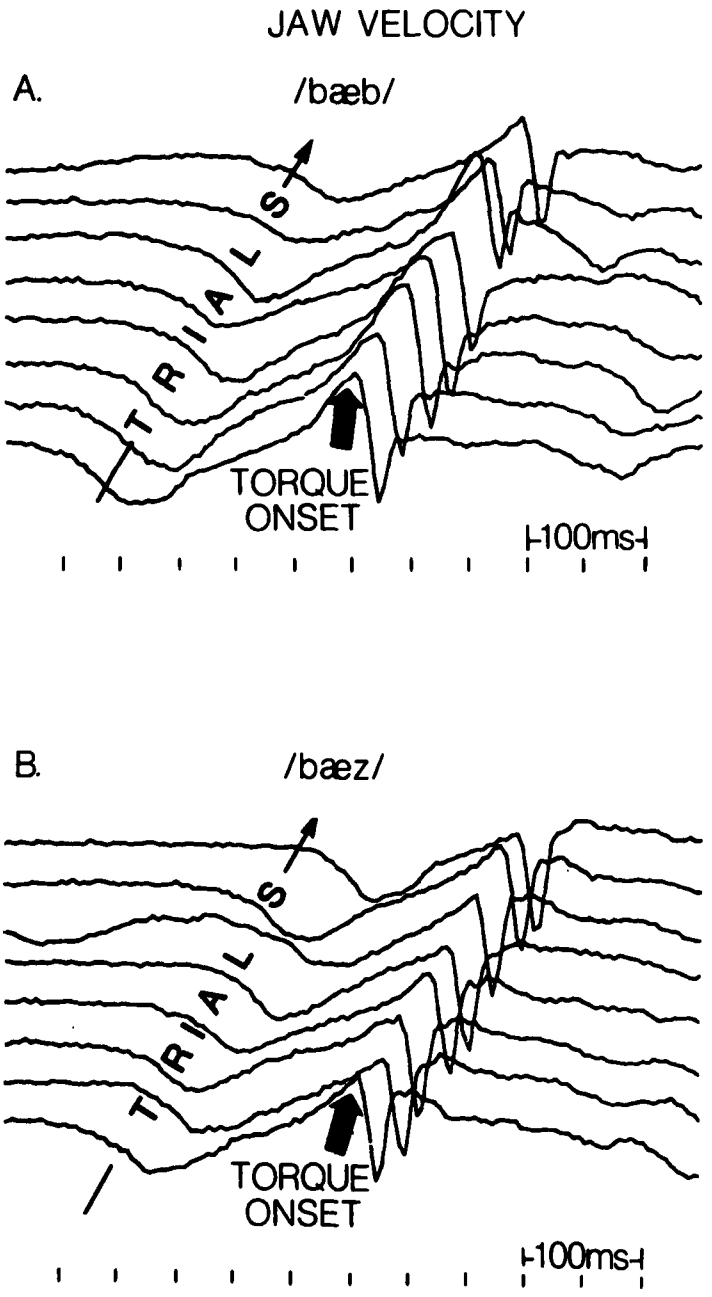


Figure 2. The consistent reaction of the jaw to a constant force load (5.88 Newtons, 1.5 s) applied during closure for the final consonant in /bæb/ and /bæz/. Velocity changes direction abruptly in response to torque. The traces are raw data and represent the first eight of a set of ten perturbation trials presented randomly in a sequence of 40 trials. The remaining two traces were very similar but are not shown because of a graphics display limitation.

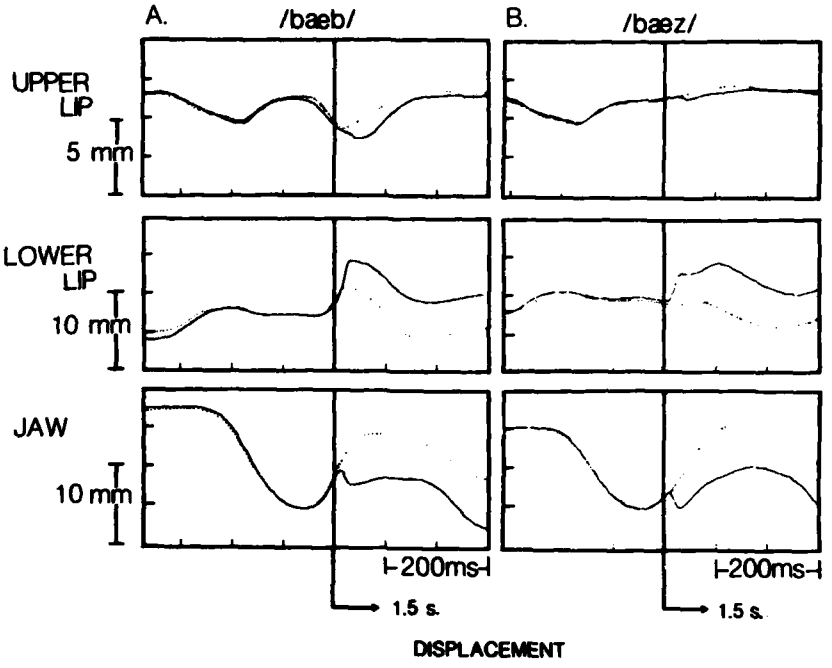


Figure 3. A and B. Upper lip, lower lip (with jaw movement contribution subtracted out) and jaw displacement for the utterances /bæb/ and /bæz/. Each trace represents the average of 10 tokens for perturbed (solid line) and control (dotted line) conditions. The vertical line in each window marks the onset of torque to the jaw. For illustration purposes, the two conditions have been overlaid by temporally sliding the control condition, which does not have a torque line-up point, relative to the perturbed condition, which does, taking the jaw as a reference point.

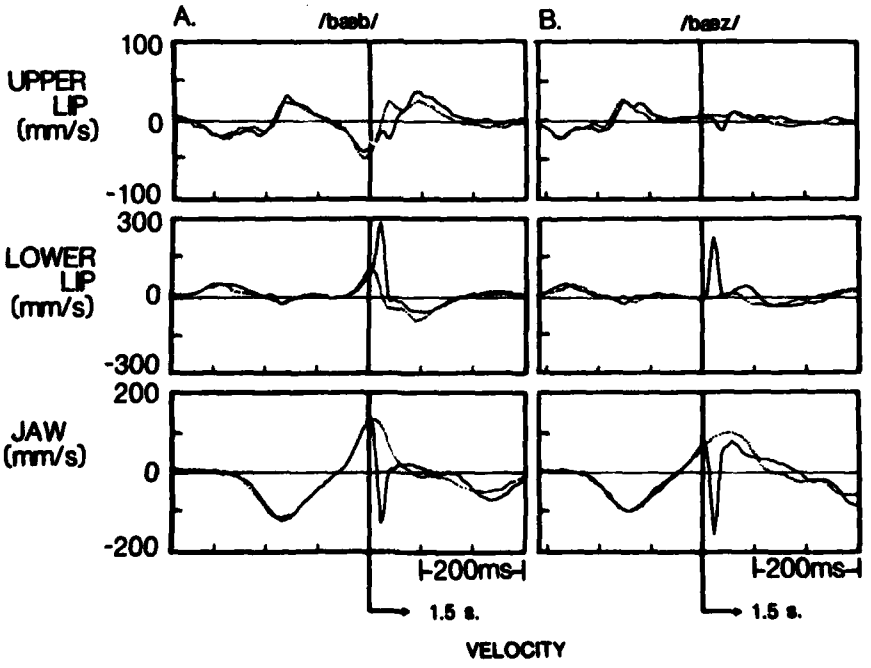


Figure 4. A and B. The articulator velocity profiles for the same data shown in Figure 3 with the same plotting conventions.

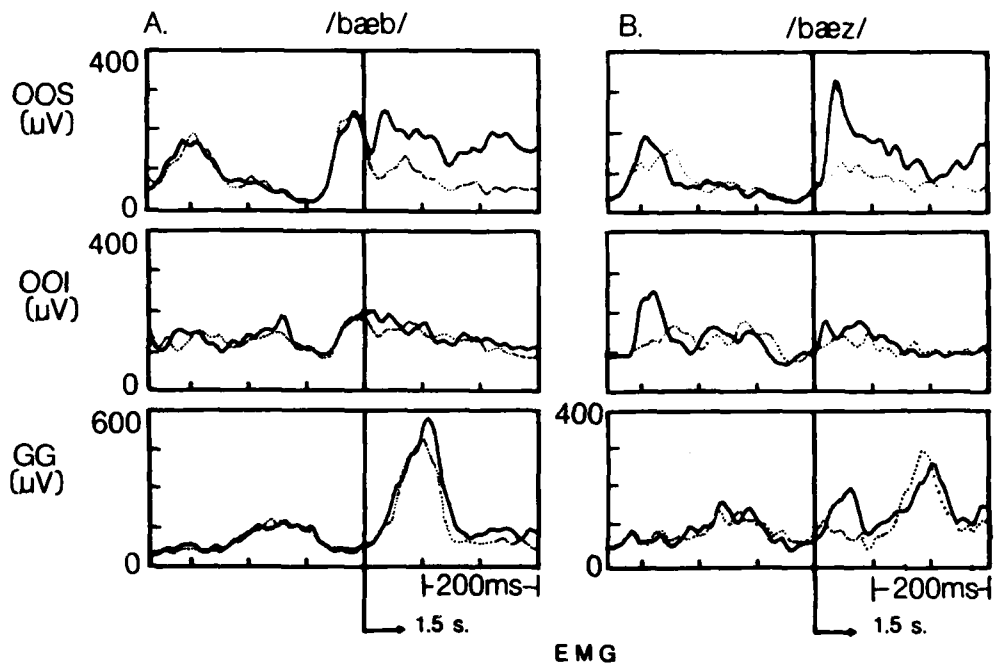


Figure 5. A and B. Average rectified electromyographic activity of upper lip (OOS), lower lip (OOI), and tongue (GG) muscles for perturbed (solid trace) and control (dotted line) conditions.

in the upper lip rather than active trajectory control. Further research is needed to evaluate these possibilities.

In contrast to the upper lip kinematics, the lower lip exhibits compensatory movement behavior in both /bæb/ and /bæz/ utterances (Figures 3 and 4). Examination of displacement and velocity profiles reveals a rapid increase in lip kinematic values when the jaw is perturbed. The near-immediate and highly consistent response of the lower lip to perturbation is shown for individual tokens in Figure 6. The onset delay of the increase in lower lip velocity--seen as an inflection point in the closing gesture for /bæb/ and as a sharp velocity spike in /bæz/--is on the order of 5 to 10 ms. As an interesting aside, the trajectory difference between the lower lip in /bæb/ and /bæz/ before perturbation suggests that the lower lip is not involved ordinarily in producing /z/ but is involved in /b/ production (see also averaged data in Figures 3 and 4).

The almost immediate response of the lower lip to jaw loading and the fact that there are no significant increases in OOI activity (Figure 5, middle row) for either utterance indicate that the lower lip perturbation response is a passive mechanical effect that arises when jaw motion is abruptly halted. In addition, the highly stereotypic lower lip reaction to jaw perturbation contrasts with other perturbation studies in speech that show considerable trial-to-trial variability in articulator movements. For example, in response to a brief perturbation applied to the lower lip, Abbs and Gracco (1983; in press) find reciprocal trade-offs in amplitude between upper and lower lip

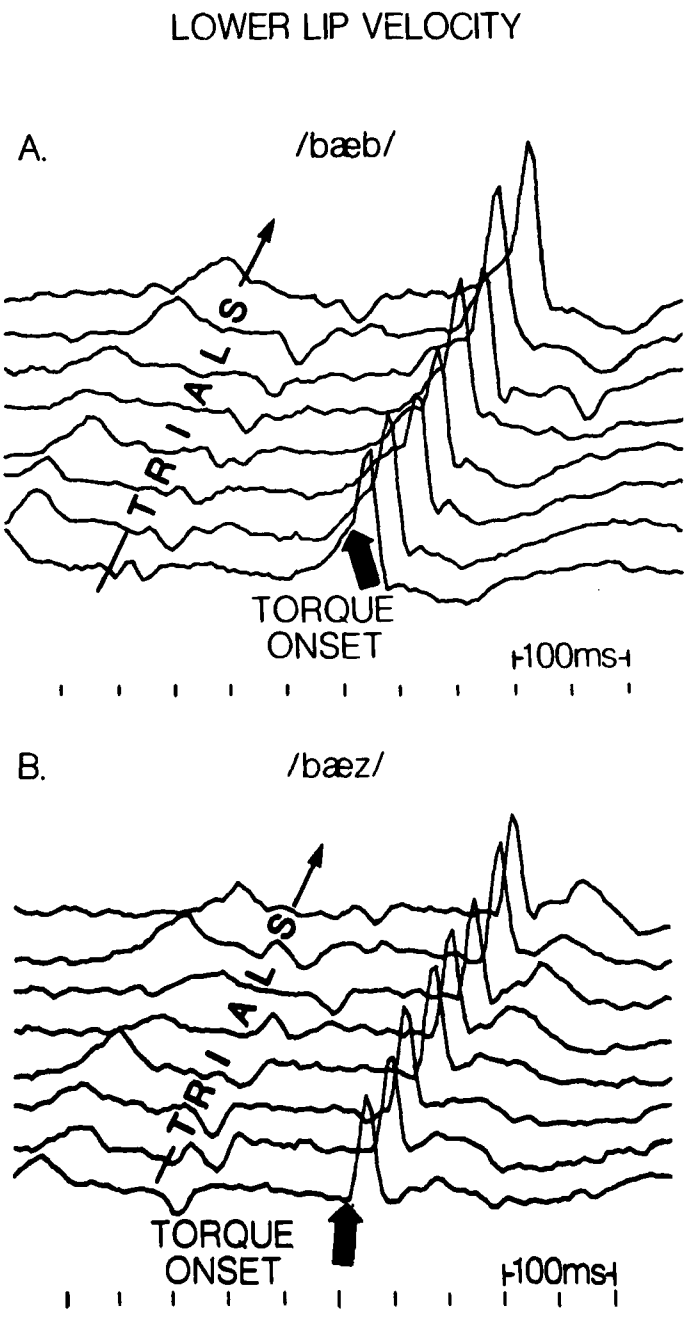


Figure 6. A and B. The very rapid and consistent lower lip reaction, seen as an inflection in the velocity trace, to perturbations of the jaw for /bæb/ and /bæz/. The plotting convention is identical to that shown in Figure 2.

movements as well as in associated muscle activity. In so-called "active compensation," different (but systematic) magnitudes of movement and EMG activity in coupled articulators appear to be the rule (see also Hughes & Abbs, 1976). The stereotypy evident in the present lower lip data, however, is more indicative of a passive shearing of the lower lip from the jaw, arising as a consequence of the momentum created by halting jaw motion.

One important feature of the lip closure response to perturbation should not be overlooked, namely, that the lips do not meet at the same point in space as they do in control conditions. In Figure 3A for example, the amplified response of the lower lip alone (solid line) does not mean that the lower lip is more elevated in perturbed than control conditions. In fact, the opposite is true because the increase in lower lip displacement is smaller than the decrease in jaw height created by loading. Thus, not only is the upper lip lower in space in perturbed relative to control conditions, but the lower lip is also, $t(18) = 3.20$, $p < .01$. What seems important here is that closure, not some spatial target, is achieved, (cf. MacNeilage, 1970, 1980, for a discussion of the status of target theories in speech).

The passive reaction of the lower lip contrasts with the active compensation to jaw loading evident in tongue muscle activity for /bæz/. When EMG responses from genioglossus are aligned and averaged with respect to the onset of /z/ frication, the increased amplitude in perturbed trials relative to control trials is highly significant, $t(18) = 7.76$, $p < .001$. Again, like the lips in /bæb/, the EMG response in /bæz/ is time-locked to the application of torque (see Figure 5B) and occurs remarkably quickly (range 20-30 ms). No such differences in tongue muscle activity occur for /baeb/, $t(18) = .88$, $p > .10$.[5]

The pattern of reactions to perturbations of the same magnitude but of much shorter duration (50 ms) was similar in some respects to those discussed above but with some marked differences. Figures 7 and 8 display the kinematic variables of displacement and velocity for each articulator and Figure 9 shows corresponding EMG data. One difference that is immediately apparent is that the articulators for both /bæb/ and /bæz/ quickly return to their normal trajectories following the offset of the perturbation (compare Figures 3 and 4 with Figures 7 and 8). In fact by the time closure is achieved, there are no significant displacement differences between perturbed and control conditions in the upper lip for /bæb/, $t(18) = 0.1$, $p > .1$. Differences in the amplitude of muscle activity in the tongue for /bæz/ come close to, but miss, significance, $t(18) = 1.84$, $p > .05$.

This homeorhetic property of the articulatory trajectories (i.e., a tendency to return to a "preferred" trajectory) has been observed before in studies of human finger (e.g., Kelso & Holt, 1980) and monkey arm movements (cf. Bizzi, Chapple, & Hogan, 1982) and has led to the proposal that trajectory is an actively controlled variable (Bizzi et al., 1982). However, the present data display lightly damped spring-like behavior; the return to a normal jaw trajectory, for example, is preceded by an overshoot response. Thus, homeorhesis may arise as a consequence of the behavior of a dynamic system and need not require the assumption of active trajectory control.

In summary, though the present findings are preliminary they are nevertheless consistent with coordinative structure theory, particularly when recent work on speech and other motor activities is also considered. For

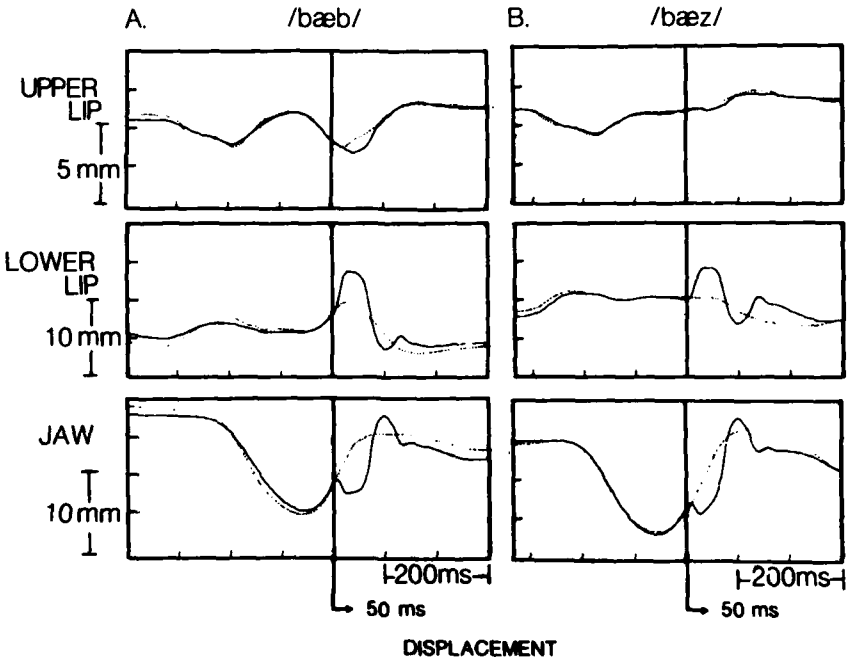


Figure 7. A and B. Upper lip, lower lip (with jaw movement contribution subtracted out) and jaw displacement for the utterances /bæb/ and /bæz/. Each trace represents the average of 10 tokens for perturbed (*solid line*) and *control* (dotted line) conditions. The vertical line in each window marks the onset of torque to the jaw. In this case a torque of 5.88 N. is applied for only 50 ms.

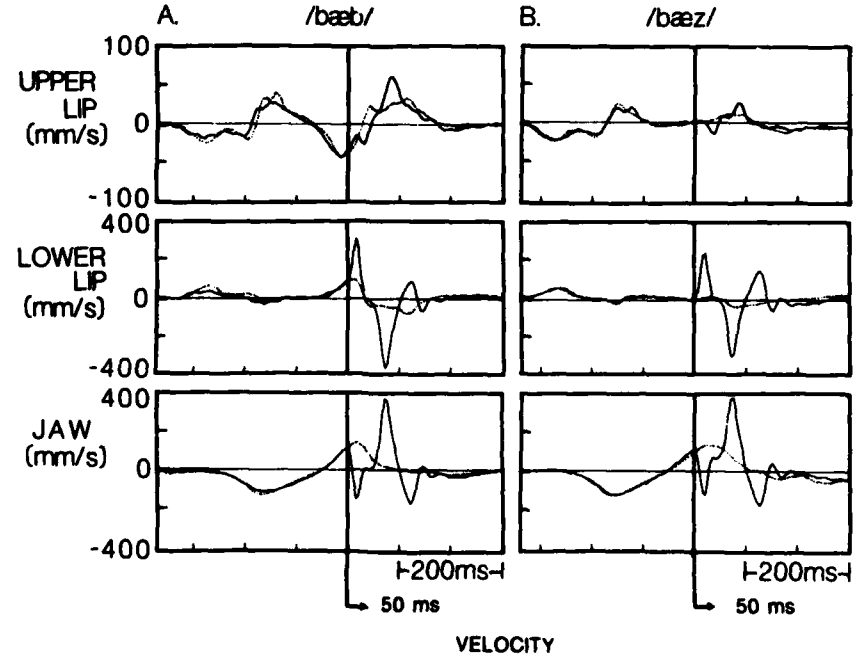
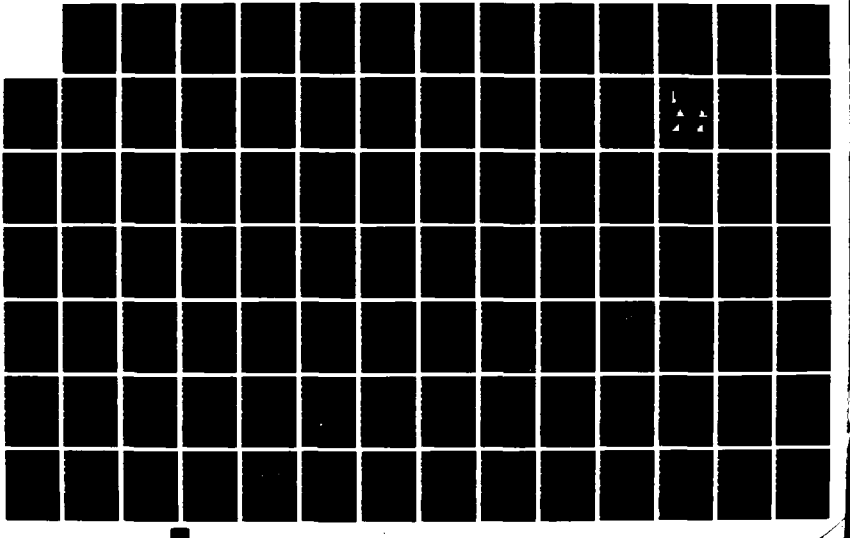
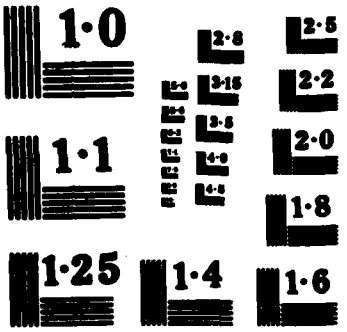


Figure 8. A and B. Corresponding articulator velocity profiles for the displacement data shown in Figure 7.

AD-A145 585 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS 2/3
AND PROGRESS OF S..(U) HASKINS LABS INC NEW HAVEN CT
A M LIBERMAN AUG 84 SR-77/78(1984) N00014-83-K-0083
UNCLASSIFIED F/G 17/2 NL




1·0
1·1
1·25
1·4
1·6
1·8
2·0
2·2
2·5

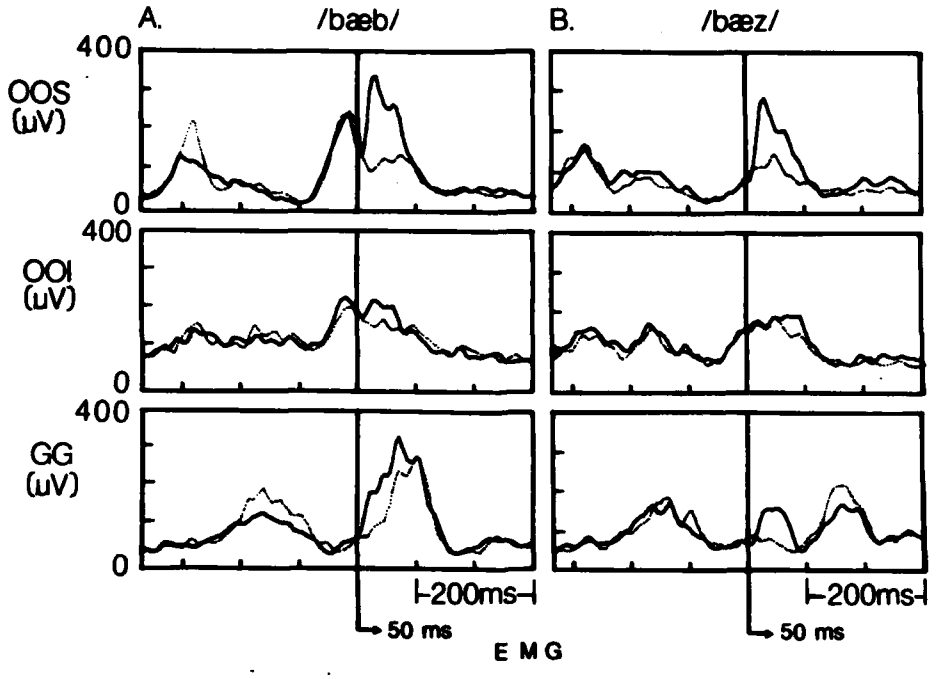


Figure 9. EMG profiles corresponding to kinematic data for briefly perturbed (solid lines) and control trials. Each trace is the average of 10 tokens.

example, the highly flexible character of the EMG and kinematic patterns observed in Experiments 1 and 2 share a likeness to recent studies of cat locomotion in which adaptive reactions are also evident (cf. Forssberg, 1982, for review). For instance, when light touch or a weak electrical shock is applied to the paw during the flexion phase of the cycle, an abrupt withdrawal response occurs as if the cat were trying to lift its leg over an obstacle. When the same stimulus is applied during the stance phase of the cycle, the flexion response (which would make the animal fall over) is inhibited, and the cat responds with added extension (cf. Forssberg, Grillner, & Rossignol, 1975). The so-called "stumble corrective reaction" is present in intact and spinal animals and, like the forms of interarticular cooperation we have observed, occurs remarkably quickly. The earliest flexor burst in response to a tactile stimulus applied during the swing phase, for example, occurs with a latency of 10 ms. Just as these reactions are non-stereotypic and functionally suited to the requirements of locomotion, so the patterns obtained in our experiments appear to be flexibly tailored to meet phonetic requirements.

In a final experiment we attempt to converge on the task-specific nature of coordinative structures by asking, in a manner akin to the research discussed above, whether the cooperative behavior among articulators is sensitive to the phase of motion during which an unexpected perturbation is applied. For example, does perturbing the jaw during the opening phase of the utterance /bæb/, induce a remote reaction in the upper lip? Since the upper lip is minimally (if at all) involved in the opening, vowel-producing phase, we would not expect to see a remote response in that phase unless the system were

rigidly coupled. However, in the closing phase (i.e., the transition out of the vowel into the final consonant) where the upper lip is actively involved in the closing gesture, the upper lip should respond to a sudden lowering of the jaw and lower lip. In addition to the question of remote reactions, we wanted to examine possible phase dependent responses in the structures local to the perturbation, namely, the lower lip and the jaw itself.

Experiment 3

Subject, Materials, and Procedures

One subject, an adult male who was not one of the authors, and who had never participated in a perturbation study, took part in this experiment (see footnote 3). The speech sample contained two utterance types "/bæb/ again" and "/bæp/ again." Eighty trials of each utterance were performed in a single block, for a total of 160 trials. In each block, 12.5% of the trials were perturbed during the opening phase of jaw motion, and 12.5% on the closing phase. The jaw was perturbed at the same predetermined position in both phases of the motion. As before, a constant force load of 5.88 Newtons and lasting 1.5 s was delivered to the jaw via a torque motor attached to a custom-made dental prosthesis. Between perturbations, the motor exerted a 30 g tracking force that did not perceptibly impede or alter normal articulation.

Once again, jaw and upper and lower lip movements were optically tracked using a modified SELSPOT system. In addition, EMG potentials from OOS and OOI were obtained from noninvasive surface (paint-on) electrodes. It is important to note that the subject knew neither which trials would be perturbed, nor the phase of jaw motion that would be loaded. An additional level of uncertainty was present, therefore, in this experiment. Movement and EMG data, and the audio signal were recorded for later off-line processing.

Results and Discussion

The following analysis of the movement trajectories is based largely on differences in peak articulator positions between perturbed and control trials for opening and closing phases of the respective gestures. First we show that the load systematically influenced jaw motion as intended. Figure 10 shows four pairs of jaw movement trajectories, corresponding to the four conditions examined. Each pair represents the averaged trajectories for all the perturbed and control trials belonging to that loading phase and phonetic context. During the opening phase of jaw movement, the perturbed trajectories, denoted by the heavier line, rapidly diverge downward after load onset. At the point of maximum opening for the vowel, they are much lower, $t(14) = 4.63$ and $t(17) = 4.59$, $ps < .001$, for /bæb/ and /bæp/, respectively.[6] Note also that the jaw trajectories are still lower at the point of peak raising for the final consonant, $t(14) = 5.21$ (/bæb/) and $t(17) = 4.26$ (/bæp/), $ps < .01$. This is perhaps not surprising, because the load remains on for 1.5 s. When the load is applied during the closing phase of motion, the jaw trajectories, as expected, are not different at peak jaw lowering for either /bæb/, $t(12) = -.20$ or /bæp/, $t(18) = -1.73$, $ps > .10$. Following load onset, however, the trajectories again diverge, and the loaded jaw remains much lower at stop closure in both phonetic contexts, $t(12) = 8.69$, $p < .01$ for /bæb/ and $t(18) = 5.23$, $p < .01$ for /bæp/. It is clear, therefore, that load application in both phases of the motion had the intended effect on the jaw trajectories.

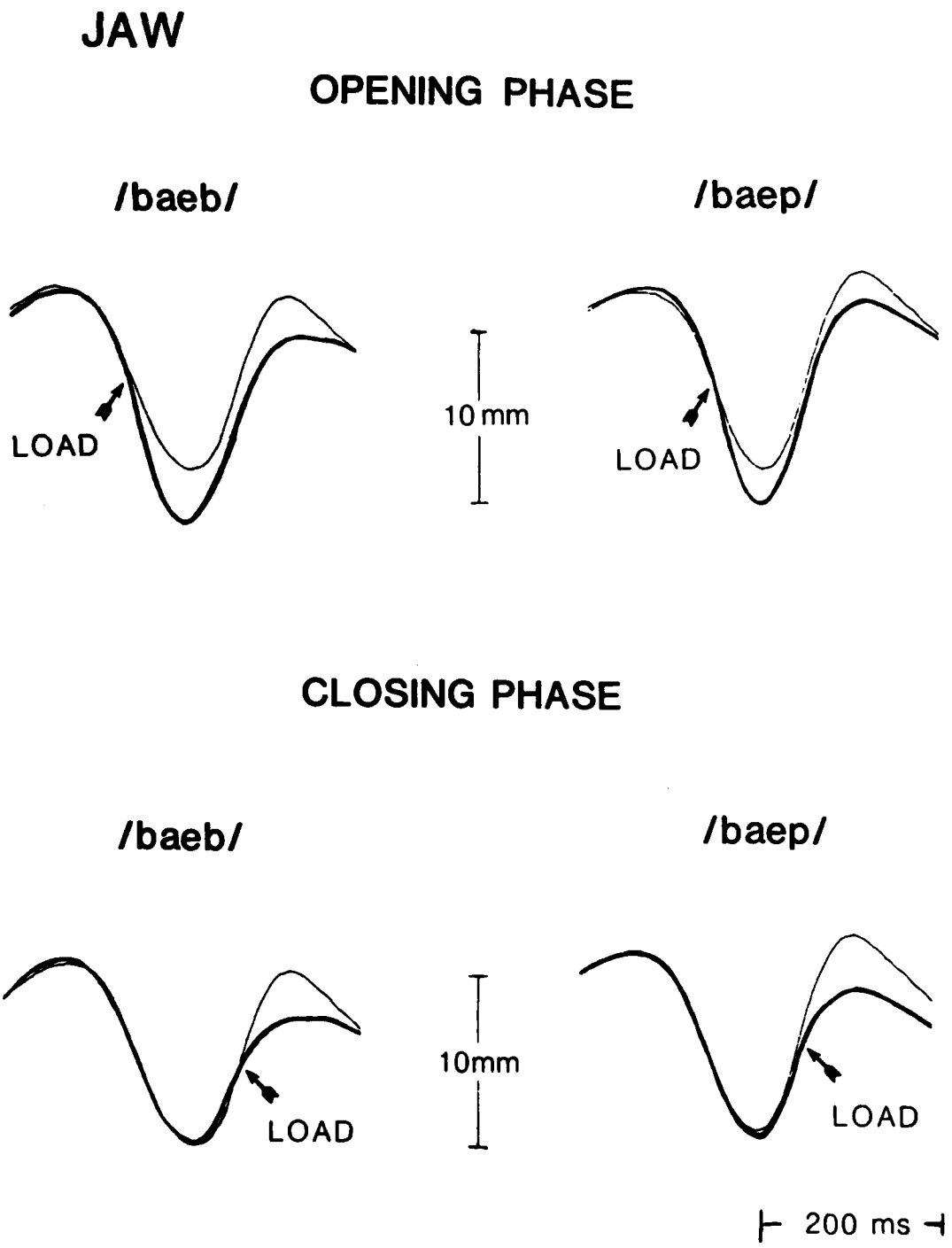


Figure 10. Four pairs of jaw movement trajectories corresponding to the four experimental conditions examined. The thin lines are the average unperturbed, control trials. Thick lines represent the mean perturbed trajectories.

In Figures 11 and 12, we show the extent to which "local" reactions occur in the lower lip in response to jaw perturbation for the utterances /bæb/ (Figure 11) and /bæp/ (Figure 12). In the figures, lower lip position is shown in absolute space as it rides the jaw (the LLJ traces) and with the jaw motion subtracted out (the LL traces). The traces along the bottom of the figure are averaged, but unsmoothed, signals for a lower lip muscle (OOI), which is active for bilabial closure. Stippled portions denote increased muscle activity in perturbed (the thicker line) relative to control trials.

Like the jaw, the lower lip-jaw complex shows a reaction to the jaw load during the opening phase of motion. Measured at maximum lowering, LLJ is perturbed downward in both /bæb/, $t(14) = 6.03$ and /bæp/, $t(17) = 5.96$, $ps < .01$. Again, since the load remains on, the lower lip-jaw combination remains lower at the point of peak closure on perturbed trials, $t(14) = 3.71$, $p < .01$ (/bæb/) and $t(17) = 4.75$, $p < .01$ (/bæp/). When the jaw is loaded during the closing phase of motion, we see a difference between perturbed and control LLJ traces only at the point of peak closure, $t(12) = 6.08$, $p < .01$ for /bæb/ and $t(18) = 5.38$, $p < .01$, for /bæp/. As expected, the trajectories are not significantly different at peak lowering, i.e., before the load is applied, $t(12) = -.47$ and $t(18) = -1.55$, $ps > .10$ for /bæb/ and /bæp/, respectively.

In Figures 11 and 12 we show also the lower lip alone (LL) responses to perturbation in the opening phase. Independently of jaw lowering, the lip traces diverge rapidly after load onset and are reliably lower at peak opening for the vowel after jaw loading in both /bæb/, $t(14) = 5.55$ and /bæp/, $t(17) = 6.00$, $ps < .01$. A marked increase in orbicularis inferior activity accompanies the lower lip response. A conservative estimate of the mean latency in OOI is 20 ms, with a 15-35 ms range. Although the mean lower lip position (relative to control) is not as high at closure in conditions when the jaw is loaded during the opening phase, the effect is highly variable and nonsignificant, $t(14) = -1.06$, $p > .10$ for /bæb/ and $t(17) = -1.31$, $p > .10$ for /bæp/.

On the right hand side of Figures 11 and 12 is shown the average lower lip response to perturbations applied during the closing phase of jaw motion. The peak closure displacements are not different between perturbed and control trials for either /bæb/, $t(12) = -1.24$, $p > .10$ or /bæp/, $t(17) = .53$, $p > .10$, suggesting that the lower lip has completely compensated for the lower jaw position. Again, there is a noticeable OOI reaction some 30 ms on the average after load onset, although this may in part reflect overall stiffening of the lower lip (note the generally elevated posture of the lower lip after peak closure has occurred). As expected, the lip trajectories are not different prior to load onset, that is, at peak lower lip depression, $t(12) = -.79$, $p > .10$ for /bæb/ and $t(18) = .86$, $p > .10$ for /bæp/.

Local movement and EMG reactions occur in response to jaw perturbations that are introduced in both opening and closing phases of the gestures. The very pronounced OOI activity when the load occurs during the opening phase of jaw motion may be indicative of the upcoming requirement of lip closure. Since the mean lower lip position (independent of jaw movement) is lower as a result of the perturbation, it must move further and more rapidly to contribute to bilabial closure. Hence an increase in muscle activity is not surprising. The active changes in lower lip muscle activity in this subject contrast with the passive "shearing" effects exhibited by a different subject in Experiment 2 (and possibly in Experiment 1 as well). Note that the form of

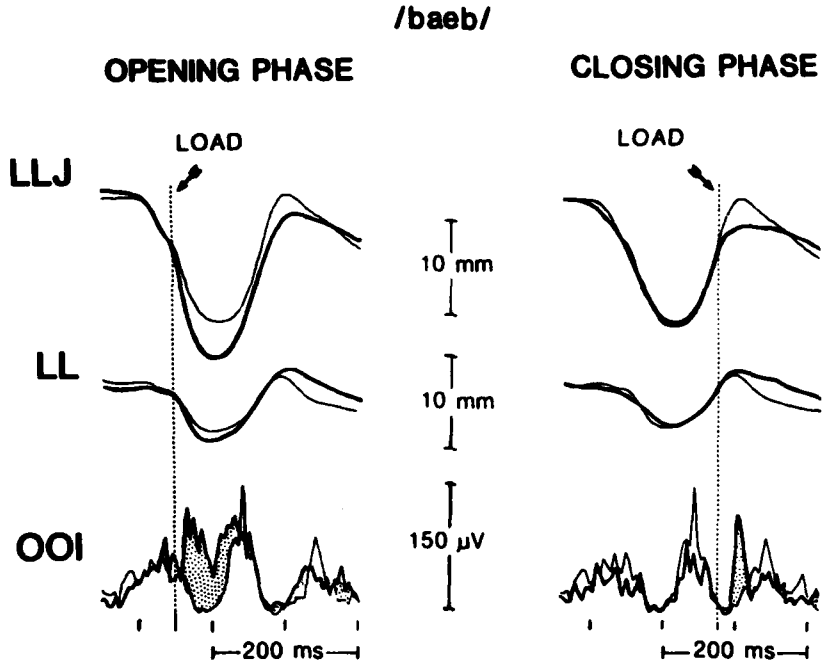


Figure 11. Average lower lip plus jaw (LLJ) and lower lip alone (LL) trajectories for the utterance /bæb/ under perturbed (thick line) and control conditions. OOI is the rectified and averaged, but unsmoothed electromyographic response of a lower lip raising muscle, orbicularis oris inferior. The thicker line denotes perturbed responses.

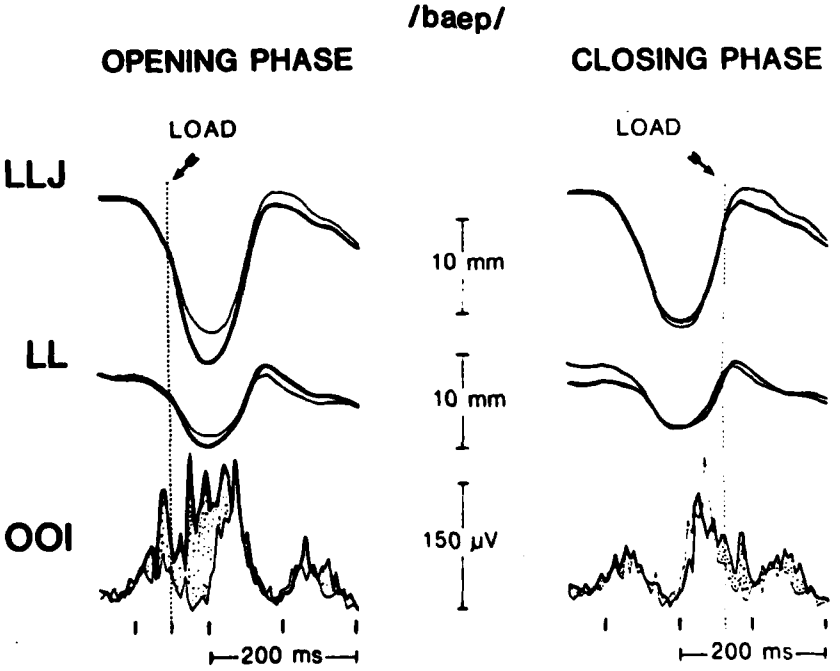


Figure 12. Average lower lip plus jaw (LLJ) and lower lip alone (LL) trajectories for the utterance /bæp/ under perturbed (thick line) and control conditions. OOI is the rectified and averaged, but unsmoothed electromyographic response of a lower lip raising muscle, orbicularis oris inferior. The thicker line denotes perturbed responses.

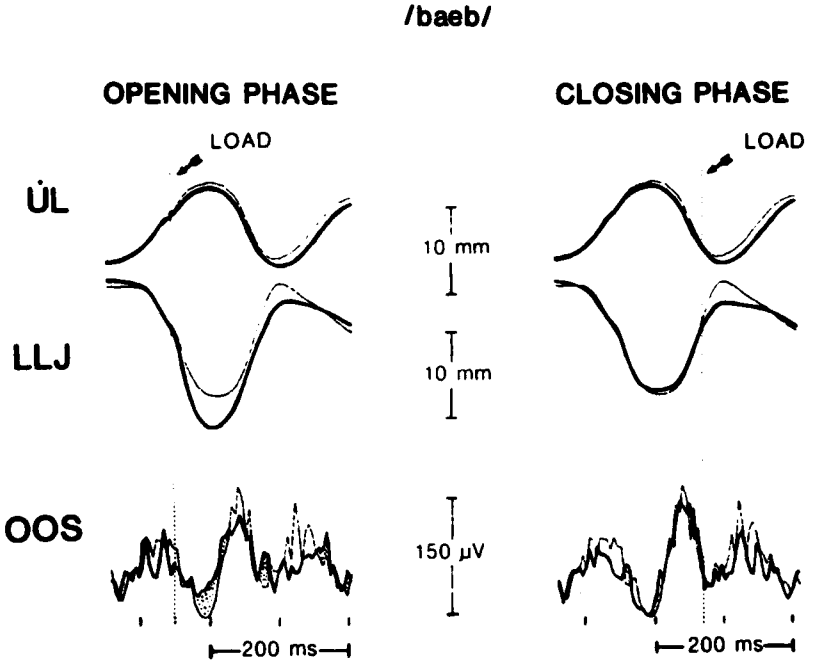


Figure 13. Average upper lip (UL) and lower lip plus jaw (LLJ) trajectories for the utterance /bæb/ under perturbed (thick line) and control conditions. OOS is the rectified and averaged, but unsmoothed, EMG response of an upper lip lowering muscle, orbicularis superior. The thicker line denotes perturbed responses.

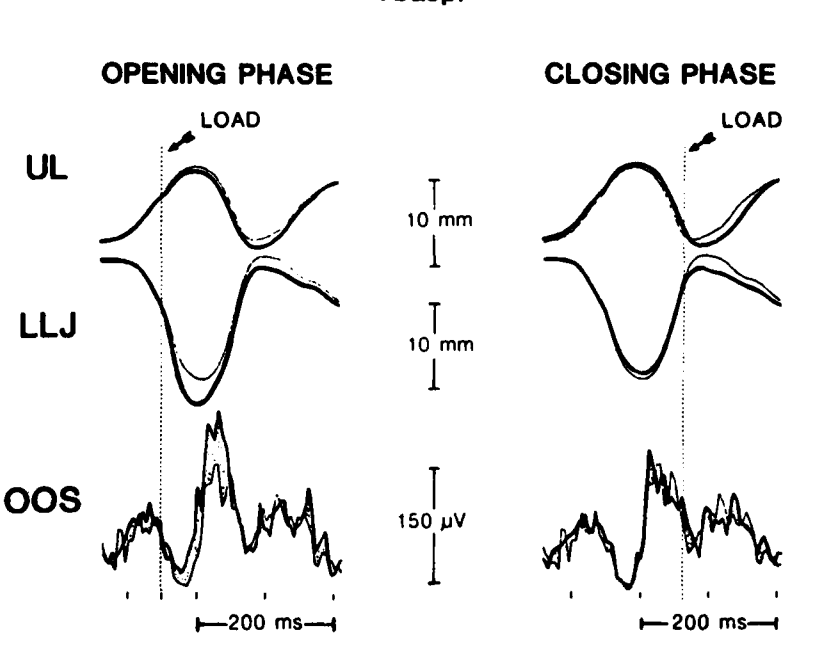


Figure 14. Average upper lip (UL) and lower lip plus jaw (LLJ) trajectories for the utterance /bæp/ under perturbed (thick line) and control conditions. OOS is the rectified and averaged, but unsmoothed, EMG response of an upper lip lowering muscle, orbicularis superior. The thicker line denotes perturbed responses.

the jaw trajectories in the same phonetic context (/bæb/) is also dramatically different between subjects. For the first subject the jaw was essentially halted by a load applied during the raising trajectory (see Figure 3). For the subject in this experiment, the load had more of a resistive effect on the jaw trajectory. These between-subject differences in jaw trajectory in reaction to a load may influence the extent to which a structure linked to the jaw (the lower lip) actively participates. A sudden halting of the jaw may cause a shearing response in the lower lip, whereas a reduction in the magnitude of the load or a stronger jaw reaction to the load may be associated with a more active neuromuscular response in locally linked articulators. A systematic manipulation of load magnitude could help resolve this question.

Although we do not expect the patterns of cooperation among articulators to be identical among subjects, we do predict (provided anatomical limitations have not been violated) that the integrity of the phonetic act will be preserved. What then of phase-dependent remote effects? In Figures 13 and 14 we display the upper lip movement and EMG traces for perturbed and control trials of /bæb/ (Figure 13) and /bæp/ (Figure 14). To aid comparison, the lower lip plus jaw trajectories are also shown. When the perturbation was applied during the opening phase, the upper lip trajectories were variable and no different from control when measured at the peak raising point, $t(14) = 1.45$, $p > .10$ for /bæb/ and $t(17) = 1.70$, $p > .10$ for /bæp/. However, in opening phase perturbation trials, the upper lip does lower further on perturbed as compared to control trials when lip position is measured at peak closure, $t(14) = 3.65$, $p < .01$ (/bæb/) and $t(17) = 3.51$, $p < .01$. Presumably this occurs to accommodate the reduction in lower lip-jaw height.

When the load was applied during closure, there was again a significant upper lip lowering response for both /bæb/, $t(12) = 2.77$, $p < .01$ and /bæp/, $t(18) = 2.68$, $p < .02$, but no differences earlier in the trajectory at the point of the peak raising movement, $t(12) = 1.22$ and $t(18) = -1.32$, $ps > .10$ for /bæb/ and /bæb/, respectively.

In general, though the upper lip muscle recordings are good, clear differences between perturbed and control trials were not readily discernible in either timing or magnitude. For this subject, at least, OOS muscle activation may be sufficient to generate upper lip motion until a collision with the lower lip occurs. In short, there may be no necessary requirement for a finely modulated EMG response in upper lip since bilabial consonants are characterized by fixed boundary conditions.

General Discussion

Even simple speech gestures involve cooperation among very many degrees of freedom operating at respiratory, laryngeal, and supralaryngeal levels. Bernstein (1967) hypothesized that rather than controlling each degree of freedom separately, the central nervous system collects multiple degrees of freedom together into functional synergies or coordinative structures that then behave, from the perspective of control, as a single unit. The present research addresses Bernstein's hypothesis in an effort to identify and analyze coordinative structures in speech. In this regard, it contrasts with much other work on motor control whose focus is restricted to actions of a single joint (see Stein, 1982, for many examples).

The hallmark of a coordinative structure as we define it (see also Boylls, 1975; Fowler, 1977; Kelso & Holt, 1980; Kelso & Saltzman, 1982; Kelso et al., 1979; Kugler, Kelso, & Turvey, 1980; Nashner et al., 1979; Turvey, 1977) is the temporary marshalling of many degrees of freedom into a task-specific, functional unit. This definition should not be confused with the traditional, reflex-based usage of synergy elaborated, for example, by Easton (1972). As Szentagothai and Arbib (1974) have pointed out, such use of the term "...is too restrictive to capture the concepts" (p. 165). Partly in response to these authors' request for "...a redefinition of synergies to revitalize motor systems research" (Szentagothai & Arbib, 1974, p. 165) we have provided a recent elaboration of coordinative structures in terms of their neurophysiological and behavioral manifestations (Kelso, Tuller, & Harris, 1981/1983; Kelso & Tuller, 1983/1984).

The task-specificity hypothesized by coordinative structure theory is supported in the present experiments. For both /b/ and /z/, rapid and highly distinctive patterns of the upper lip, lower lip, and tongue occurred in response to unexpected jaw loadings so that the desired sound was produced. In all cases, the adjustments, though varied, were such as to preserve the integrity of the phonetic act. For example, for /z/ frication in Experiments 1 and 2, there was no detectable upper lip movement. But, since the jaw was much lower than usual, highly amplified tongue muscle activity, necessary to obtain an appropriate alveolar position for fricative production was observed. Like the lips in /bæb/, the tongue in /bæz/ responded remarkably quickly on the very first perturbation trial and again with no slurring or distortion perceptible to a listener. As in recent studies of bite-block speech (akin to speaking with a pipe in one's mouth), in which sensory information was drastically reduced by anesthetization of oral structures combined with auditory masking, we found no evidence of any short-term "learning" (cf. Kelso & Tuller, 1983). Articulatory "compensation" was achieved, therefore, with little or no practice.

The coordinative structure account applies equally well to disruptions that are static and anticipated (like the bite-block experiments) and those that are time-varying and unanticipated. Adjustment to either type of perturbation is a predictable outcome of an ensemble whose constituent muscles function cooperatively as a single unit. If the operation of certain variables is fixed, as in the bite-block case, or unexpectedly disturbed as a result of on-line perturbation, functionally linked variables will preserve the synergistic constraint. As we have emphasized before (Kelso & Tuller, 1983; see also Abbs & Gracco, 1983) so-called "compensation" is characteristic of the speech system's normal mode of operation. For example, in a study of respiratory function during speech, Hixon, Mead, and Goldman (1976) found that the relative contributions of thorax and abdomen movements adjust in order to preserve subglottal pressure level across large postural changes (e.g., lying versus standing). Similarly, Sussman, MacNeilage, and Hanson (1973) in a study of lip and jaw movements in a variety of vowel-consonant-vowel (VCV) triads observed that jaw elevation at consonant closure was directly proportional to the height of the following vowel. Thus, in order to occlude the vocal tract for /p/ in /æpæ/ versus /æpi/ the lips must "compensate" differentially to accommodate different jaw positions. Both of these studies suggest task-specific cooperation in naturally occurring situations.

One account of multimovement adjustments to unanticipated disruptions posits a closed-loop peripheral feedback mechanism (cf. Abbs, 1979; Folkins & Abbs, 1975). As we have pointed out, however (Fowler & Turvey, 1978, 1980; Kelso, 1981; Kelso & Tuller, 1983), a closed-loop system, though capable in theory of detecting and correcting "errors" in the perturbed structure, has no mechanism for producing adaptive movements in remote and non-biomechanically linked articulators. Because of this limitation, Abbs and Gracco (1983) have recently proposed an "open-loop adjustment process" to account for upper lip changes to lower lip perturbations "...based upon a pre-established sensorimotor translation between lower-lip afferent signals and upper lip motor actions." This notion is similar to the predictive, feedforward processes hypothesized by Ito (1975) for vestibular-ocular interactions during eye-head movement, and elaborated more recently by Houk and Rymer (1981). Viable though feedforward may be, it is nevertheless difficult to envisage how--without the concept of coordinative structure--all the computation could be pre-established in such a way that the lips, jaw, and tongue (not to mention other possible articulators not observed in these experiments) perform precisely those movements that meet the speaker's objective. The problem is exacerbated when unexpected challenges are introduced whose dimensions (e.g., magnitude, duration, site) are potentially manifold. However, although the particular neural processes involved await clarification, a central conclusion of Abbs and colleagues' work, that the "...nervous system prioritizes acoustically and aerodynamically significant multiaction gestures over individual movements and muscle actions.." and that "...these sensorimotor capabilities relieve the nervous system of having to prespecify the motor details" (Abbs, in press) has much in common with the concept of coordinative structure advocated here and elsewhere.

The results of the third experiment provide further evidence for a task-specific coordinative structure style of motor control. Remote responses in upper lip were found to be phase-dependent; that is, they occurred only when they were functionally appropriate. Similar task-dependent forms of articulator cooperation have been observed in recent studies of posture in humans (e.g., Cordo & Nashner, 1982; Marsden, Merton, & Morton, 1981, 1983). For example, Marsden et al. (1983) applied a small perturbation to the thumb of a standing subject as he was performing a thumb tracking task, and observed reactions in muscles remote from the prime mover (e.g., in pectoralis major; in the triceps of the opposite limb when it gripped a table top; in the opposite thumb when it served to stabilize motion, etc.). These distant reactions were very rapid (e.g., 40 ms in pectoralis), sometimes faster than the local autogenetic response in the structure perturbed. Though exquisitely sensitive they are not caused by length changes in the postural muscles themselves. Perturbations of only 7.5 g to the thumb or wrist, often not even detected by the subject, were associated with brisk, distant reactions. Finally and importantly, distant reactions occurred only when they performed a useful function and they were flexibly tuned to that function. Postural responses in triceps disappeared if the hand was not exerting a firm grip on the object. If, instead of holding a table top, the non-tracking hand held a cup of tea, the responses in triceps reversed, which is exactly what they have to do to prevent the tea from spilling. Marsden et al. (1983) conclude that these rapid, remote effects "...constitute a distinct and apparently new, class of motor reaction" (p. 645) that has caused them to abandon an account based on stretch reflexes. As the previous discussion indicates, however, similar phenomena have been present (although perhaps not sufficiently recognized until recently) in the speech literature as well.

To summarize, though the speed of the adaptive reactions observed in our experiments could be described as reflexive, their mutability speaks against any fixed reflex connections or rigidly constructed servomechanisms. Thus, the system we are dealing with appears to be "softly" assembled and flexible in function; not machine-like and rigid (Iberall, 1978; see also Abbs & Gracco, 1983). Similarly, it is extremely doubtful that the articulatory patterns observed here in response to jaw loading at different phases of motion and in different phonetic contexts are programmed completely in advance.

The present data, preliminary though they are, suggest nevertheless that the mode of operation of the speech system is intrinsically task-oriented, and that both rapid local and remote articulatory contributions are involved in the implementation of cooperative action. But most importantly, the adjustments appear to reflect a synergistic organization among articulators that is tailored to the requirements of the spoken act. As Bernstein (1967, p. 69) intimated:

> Movements react to one single detail with changes in a whole series of others that are sometimes very far from the former both in space and time ... In this way movements are not chains of details but structures which are differentiated into details ....

Or consider Dewey's (1896: cited in Fearing, 1930) remarks that the relations between sensory stimuli and motor consequences do not constitute a "fixed existence" but a "flexible function".' Herein lie kernel themes for a research program on coordinative structures that differs radically from approaches that *focus on control around a single joint*. The present work represents only a modest, but we think promising, beginning.

References

Abbs, J. H. (1979). Speech motor equivalence: The need for a multi-level control model. Proceedings of the Ninth International Congress of Phonetic Sciences, 2, 318-324.

Abbs, J. H. (in press). Invariance and variability in speech production: A distinction between linguistic intent and its neuromotor implementation. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability in speech procceses. Hillsdale, NJ: Erlbaum.

Abbs, J. H., & Cole, K. J. (1982). Consideration of bulbar and suprabulbar afferent influences upon speech motor coordination and programming. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), Speech motor control. Oxford: Pergamon Press.

Abbs, J. H., & Gracco, V. L. (1983). Sensorimotor actions in the control of multimovement speech gestures. Trends in Neuroscience, 6, 391-395.

Abbs, J. H., & Gracco, V. L. (in press). Control of complex motor gestures and orofacial muscle responses to load perturbations of the lips during speech. Journal of Neurophysiology.

Bernstein, N. A. (1967). The coordination and regulation of movements. London: Pergamon Press.

Bizzi, E., Chapple, W., & Hogan, N. (1982). Mechanical properties of muscles: Implications for motor control. Trends in Neuroscience, 5, 395-398.

Boylls, C. C. (1975). A theory of cerebellar function with applications to locomotion. II. The relation of anterior lobe climbing fiber function to locomotor behavior in the cat (COINS Technical Report 76-1). Amherst,

MA: University of Massachusetts, Department of Computer and Information Science.

Cordo, P. J., & Nashner, L. M. (1982). Properties of postural adjustments associated with rapid arm movements. Journal of Neurophysiology, 47, 287-302.

Easton, T. A. (1972). On the normal use of reflexes. American Scientist, 60, 591-599.

Evarts, E. V. (1982). Analogies between central motor programs for speech and for limb movements. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), Speech motor control. Oxford: Pergamon Press.

Fearing, F. (1930). Reflex action. A study in the history of physiological psychology. Cambridge, MA: MIT Press (reprinted 1970).

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech: Responses to resistive loading of the jaw. Journal of Speech and Hearing Research, 18, 207-220.

Folkins, J. W., & Zimmermann, G. N. (1982). Lip and jaw interaction during speech: Responses to perturbation of lower-lip movement during bilabial closure. Journal of the Acoustical Society of America, 71, 1225-1233.

Forssberg, H. (1982). Spinal locomotion function and descending control. In B. Sjolund & A. Bjorkland (Eds.), Brainstem control of spinal mechanisms. New York: Ferström Foundation Series.

Forssberg, H., Grillner, S., & Rossignol, S. (1975). Phase dependent reflex reversal during walking in chronic spinal cats. Brain Research, 55, 247-304.

Fowler, C. A. (1977). Timing control in speech production. Bloomington, IN: Indiana Linguistics Club.

Fowler, C. A., Rubin, P., Remez, R. E., & Turvey, M. T. (1980). Implications for speech production of a general theory of action. In B. Butterworth (Ed.), Language production. New York: Academic Press.

Fowler, C. A., & Turvey, M. T. (1978). Skill acquisition: An event approach with special reference to searching for the optimum of a function of several variables. In G. Stelmach (Ed.), Information processing in motor control and learning. New York: Academic Press.

Fowler, C. A., & Turvey, M. T. (1980). Immediate compensation in bite-block speech. Phonetica, 37, 306-326.

Gelfand, I. M., Gurfinkel, V. S., Tsetlin, M. L., & Shik, M. L. (1971). Some problems in the analysis of movements. In I. M. Gelfand, V. S. Gurfinkel, S. V. Fomin, & M. Tsetlin (Eds.), Models of the structural-functional organization of certain biological systems. Cambridge, MA: MIT Press.

Greene, P. H. (1972). Problems of organization of motor systems. In R. Rosen & F. Snell (Eds.), Progress in theoretical biology. New York: Academic Press.

Greene, P. H. (1982). Why is it easy to control your arms? Journal of Motor Behavior, 14, 260-286.

Grillner, S. (1982). Possible analogies in the control of innate motor acts and the production of sound in speech. In S. Grillner, B. Lindblom, J. Lubker, & A. Persson (Eds.), Speech motor control. Oxford: Pergamon Press.

Hamlet, S. L., & Stone, M. (1978). Compensatory alveolar consonant production induced by wearing a dental prosthesis. Journal of Phonetics, 6, 227-248.

Hixon, T. J., Mead, J., & Goldman, M. D. (1976). Dynamics of chest wall during speech production: Function of the thorax, ribcage, diaphragm and abdomen. Journal of Speech and Hearing Research, 19, 297-356.

Houk, J. C., & Rymer, W. (1981). Neural control of muscle length and tension. In V. B. Brooks (Ed.), Handbook of physiology; Sec. 1: Vol. II.: Motor control, Part 1 (pp. 257-323). Bethesda, MD: American Physiological Society.

Hughes, O. M., & Abbs, J. H. (1976). Labial mandibular coordination in the production of speech: Implications for the operation of motor equivalence. Phonetica, 33, 199-221.

Iberall, A. S. (1978). Cybernetics offers a (hydrodynamic) thermodynamic view of brain activities. An alternative to reflexology. In F. Brambilla, P. K. Bridges, E. Endroczi, & C. Heusep (Eds.), Perspectives in endocrine psychobiology. New York: Wiley.

Ito, M. (1975). The control mechanisms of cerebellar motor systems. In E. V. Evarts (Ed.), Central processing of sensory input leading to motor output (pp. 293-304). Cambridge, MA: MIT Press.

Kelso, J. A. S. (1981). Contrasting perspectives on order and regulation in movement. In J. Long & A. Baddeley (Eds.), Attention and performance (IX). Hillsdale, NJ: Erlbaum.

Kelso, J. A. S., & Holt, K. G. (1980). Exploring a vibratory systems analysis of human movement production. Journal of Neurophysiology, 43, 1183-1196.

Kelso, J. A. S., & Saltzman, E. L. (1982). Motor control: Which themes do we orchestrate? The Behavioral and Brain Sciences, 5, 554-557.

Kelso, J. A. S., Southard, D. L., & Goodman, D. (1979). On the nature of human interlimb coordination. Science, 203, 1029-1031.

Kelso, J. A. S., & Tuller, B. (1983). "Compensatory articulation" under conditions of reduced afferent information: A dynamic formulation. Journal of Speech and Hearing Research, 26, 217-224.

Kelso, J. A. S., & Tuller, B. (1983/1984). A dynamical basis for action systems. Haskins Laboratories Status Report on Speech Research, SR-73, 177-216. Also in M. S. Gazzaniga (Ed.), Handbook of cognitive neuroscience. New York: Plenum.

Kelso, J. A. S., Tuller, B. H., & Harris, K. S. (1981/1983). A 'dynamic pattern' perspective on the control and coordination of movement. Haskins Laboratories Status Report on Speech Research, SR-65, 157-196. Also in P. MacNeilage (Ed.), The production of speech. New York: Springer-Verlag.

Kugler, P. N., Kelso, J. A. S., & Turvey, M. T. (1980). On the concept of coordinative structures as dissipative structures: I. Theoretical lines of convergence. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior (pp. 1-47). New York: North-Holland.

Lindblom, B., & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw, and larynx movement. Journal of the Acoustical Society of America, 50, 1166-1179.

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. Psychological Review, 77, 182-196.

MacNeilage, P. F. (1980). Distinctive properties of speech motor control. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. Amsterdam: North Holland.

Marsden, C. D., Merton, P. A., & Morton, H. B. (1981). Anticipatory postural response in the human subject. Journal of Physiology (London), 275, 47P-48P.

Marsden, C. D., Merton, P. A., & Morton, H. B. (1983). Rapid postural reactions to mechanical displacement of the hand in man. In J. E. Desmedt (Ed.), Motor control mechanisms in health and disease (pp. 645-659). New York: Raven Press.

Nashner, L. M., Woolacott, M., & Tuma, G. (1979). Organization of rapid response to postural and locomotor-like perturbations of standing man. Experimental Brain Research, 36, 463-476.

Riordan, C. J. (1977). Control of vocal-tract length in speech. Journal of the Acoustical Society of America, 62, 998-1002.

Soechting, J. E., & Lacquaniti, F. (1981). Invariant characteristics of a pointing movement in man. Journal of Neuroscience, 1, 710-720.

Stein, R. B. (1982). What muscle variables does the central nervous system control? The Behavioral and Brain Sciences, 5, 535-577.

Sussman, H. M., MacNeilage, P. F., & Hanson, R. J. (1973). Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. Journal of Speech and Hearing Research, 16, 397-420.

Szentagothai, J., & Arbib, M. A. (Eds.). (1974). Conceptual models of neural organization. Neurosciences Research Program Bulletin, 12(3).

Tuller, B., & Fitch, H. (1980). Preservation of vocal tract length in speech: A negative finding. Journal of the Acoustical Society of America, 6, 1068-1071.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), Perceiving, acting and knowing: Toward an ecological psychology. Hillsdale, NJ: Erlbaum.

Zimmermann, G., Kelso, J. A. S., & Lander, L. (1980). Articulatory behavior pre and post full-mouth tooth extraction and alveoloplasty: A cinefluorographic study. Journal of Speech and Hearing Research, 23, 630-645.

Footnotes

[1]Initially Folkins and Abbs (1975, p. 218) interpreted their data as support for online feedback processing, that is, "a lip control system that is adjusted on the basis of feedback information about the relative position of the lips and jaw." A more recent interpretation, or perhaps a redescription by Abbs and Cole (1982, p. 171) is that the data support "a feedforward, open-loop control process..." in which "...information is fed forward for making adjustments in motor commands to structures having parallel involvements." Suprabulbar pathways are hypothesized to play a mediating role.

[2]Anecdotal evidence for such tailoring is reported by Abbs and Gracco (1983), who noticed that upper lip compensation to a lower lip perturbation occurs in the utterance /aba/ but not in /afa/. Neither data nor reference citation to this finding is presented, however. Similarly, Folkins and Zimmermann (1982, p. 1232) conclude their paper on electrical stimulation of the lower lip with the suggestion that "...it may be that interactions between the lips and jaw may be different for bilabial closing, bilabial opening, labiodental closing, and lip rounding gestures." (italics ours). Again, a direct test of this hypothesis, which we conduct here, has not been made. In fact, all the dynamic perturbation studies conducted thus far have involved bilabial gestures.

[3]Some explanation is necessary about the small number of subjects and the chronological aspects of the research. Since these experiments started in late 1978 we have tried to prepare a total of four subjects for participation. In each case special dental casts were made of the upper and lower teeth, prior to constructing a titanium prosthesis for the lower jaw. Only with two subjects, however, was it possible to proceed according to plan for the following reasons. First, in order to seat the prosthesis in the mouth firmly

so that it did not come out or reverberate when a load was applied, it was necessary to have a subject who had at least one (preferably several) of the rear molars missing (see Figure 1C). Second, and relatedly, it was crucial to have sufficient clearance at the sides of the subject's mouth so that the protruding rods to the torque motor did not interfere in any way with the subject's speech. Two subjects met these criteria, though the second subject did not become available until early 1983. We tried to test him in the larger version of Experiment 1, but he was unable to withstand the insertion of fine wire electrodes into the tongue and hence could not be used to study fricative production. Because of these difficulties we can report only our efforts to provide a within-subject replication of the experiment (Experiment 2). The second subject, however, participated in Experiment 3, which did not require invasive procedures. We did not run subject 1 in the latter study because we were concerned about possible experiential factors influencing the results.

[4]Peak lip displacement can occur after closure is attained because of the elastic nature of the lips. Once the upper and lower lips touch, achieving closure, they can and usually do compress further as closure proceeds.

[5]The large burst of genioglossus activity evident in /bæb/ and also the second peak in /bæz/ is related to production of the /g/ in the carrier phrase "again." Examination of the acoustics revealed that the torque occurred closer to the onset of /b/ closure than to /z/ frication. This is reflected in the proximity of genioglossus activity to torque onset in /bæb/ relative to /bæz/.

[6]In the following analyses, there are always ten control trials to compare with the perturbed trajectories. However, because of technical difficulties (e.g., the subject making non-speech jaw movements that triggered the perturbation), there are not always ten perturbed trials. We present therefore the pooled degrees of freedom (N-2) for statistical tests, although we have performed all the tests using the adjusted degrees of freedom (after Scheffé) as well. Pooled and adjusted results are very similar; however, where they diverge we will report both.

[7]Fearing's (1930) book is a most scholarly treatment of the reflex concept in psychology and physiology. Given recent findings (see General Discussion) the book has a prophetic tone. For example, Dewey's remarks made in 1896, offer a stark contrast with Sherrington's in 1906. Sherrington on the one hand admitted that the reflex was a "likely if not probable fiction," but on the other referred to it as having "a machine-like fatality" (cited in Fearing, 1930, Chapter 16). Fearing's conclusion (pp. 313-315), in which he advocates an experimental approach that does not focus on isolable fragments of an action, but rather examines the relations among concomitant events in the integrated nervous system, is anticipatory of some, but by no means all, current work on motor control.

FORMANT INTEGRATION AND THE PERCEPTION OF NASAL VOWEL HEIGHT*

Patrice Speeter Beddor

Abstract. Research on oral vowels has shown that vowel perception involves integration of adjacent spectral components such that perceived height correlates with the center of the first region of spectral prominence or "center of gravity". This study investigated the center-of-gravity effect in nasal vowels and asked whether formant integration in vowel perception extends to the first oral formant, F1, and the first nasal formant, FN. Five nasal vowels, [ĩ ẽ æ̃ ã õ], were synthesized. For each nasal vowel, a continuum of synthetic oral vowels was generated by manipulating the frequency of F1. Five vowel sets were constructed by pairing the nasal vowel standard with each member of the corresponding oral vowel continuum; listeners selected the "best-match" pair for each set. Listeners chose the oral-nasal pairs with the same F1 frequency in vowel set i only. For e, æ, a, and o, listeners' matches depended on the relative position of F1 and FN in the nasal vowel: when FN frequency was less than F1, as in [æ̃] and [ã], the best oral match had a relatively low F1 frequency; when FN frequency exceeded F1, as in [ẽ] and [õ], the oral match had a high F1. These perceptual data indicate spectral averaging of adjacent oral and nasal vowel formants, thereby demonstrating the center-of-gravity effect in the perception of nasal vowels.

This paper reports the results of a study of the acoustic features determining perceived height in nasal vowels. Most previous research of the perception of vowel height has dealt with oral vowels. Phoneticians generally acknowledge that the perceptual dimension of height in oral vowels is inversely correlated with the frequency of the first formant, such that height perceptually lowers as first formant frequency increases (Fant, 1960; Joos, 1948; Ladefoged, 1982; Peterson & Barney, 1952). But despite this correlation, the frequency of the first formant is not the sole determinant of perceived vowel height.

---

*A shorter version of this paper was presented at the Annual Meeting of the Linguistic Society of America in Minneapolis on December 30, 1983.

Acknowledgment. This work was supported by NIH Postdoctoral Fellowship Grant NS-07196 to the author and by NIH Biomedical Research Support Grant RR-05596 to Haskins Laboratories. I am grateful to Bruno Repp and Terry Gottfried for advice on experimental design and data analysis and to Phil Rubin for programming the centroid routine used in this study. I also wish to thank Sarah Hawkins, Ignatius Mattingly, and Kathleen Houlihan for helpful comments on an earlier draft.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

Experimental evidence indicates that the first formant does not acoustically specify height in oral vowels when the frequencies of the first two vowel formants are relatively close together, as in back vowels. Studies with synthetic vowels have shown that perceived height in back vowels is determined not only by the first formant (F1), but also by the second formant (F2). In experiments where one formant vowel approximations were perceptually matched to two-formant back vowel stimuli, the frequency of the single formant was not matched to F1 or F2 of the two-formant stimulus, but was instead located between F1 and F2 (Bedrov, Chistovich, & Sheikin, 1978). Similarly, Delattre, Liberman, Cooper, and Gerstman (1952) found that reduction in F1 amplitude perceptually lowered back (but not front) vowels, while reduction in F2 amplitude perceptually raised back vowels, leading to the speculation that "the ear effectively averages two vowel formants which are close together" (1952, p. 203).

Perceptual averaging of vowel spectrum components that are relatively close in frequency is not restricted to F1 and F2, but also occurs for F2 and F3 (Bladon & Fant, 1978; Carlson, Fant, & Granström, 1975; Carlson, Granström, & Fant, 1970; Miller, 1953) as well as for the first harmonic and F1 (Carlson, Fant & Granström, 1975; Fujisaki & Kawashima, 1968; Traunmüller, 1981). A substantial body of data therefore indicates that perception of vowel quality involves calculation of a weighted mean of adjacent spectral prominences rather than merely extraction of the frequencies of the spectral peaks. That is, when two spectral prominences fall within some critical frequency range, vowel quality is determined by the "center of gravity" of the region of prominence (Chistovich & Lublinskaya, 1979; Chistovich, Sheikin, & Lublinskaya, 1979). The center-of-gravity effect disappears when the distance between spectral peaks exceeds 3.0 to 3.5 Bark (Chistovich & Lublinskaya, 1979; Syrdal & Gopal, 1983).[1]

This study extends investigation of the center of gravity effect to nasal vowels. The acoustic theory of vowel nasalization predicts that velopharyngeal coupling of the nasal tract to the main vocal tract adds pole-zero pairs and shifts formant frequencies of the transfer function of the coupled system (i.e., nasal vowel) relative to the transfer function of the uncoupled (non-nasal) system. Especially important to this study of nasal vowel height is that the main acoustic effect of nasal coupling is in the region of F1, where F1 of the non-nasal vowel is replaced in the nasal vowel by two poles and a zero (Fant, 1960; Fujimura & Lindqvist, 1971; Hamada, 1983; Stevens, Fant, & Hawkins, in press). The two poles are the first nasal formant and the first oral formant, the latter typically being shifted in frequency, with a wider bandwidth and lower amplitude than the first formant of the non-nasal vowel (Delattre, 1954; House & Stevens, 1956; Mrayati, 1975). Thus the low-frequency region of nasal vowel spectra is characterized by a relatively flat, wide distribution of acoustic energy (see Maeda, 1982). Some of these spectral properties of nasal vowels are illustrated in Figure 1 by the spectrum of a Hindi speaker's nasal [ẽ] (solid curve), superimposed on the spectrum of Hindi oral [e] (dashed curve). Note that the low-frequency spectral energy of [ẽ] is spread across two broad spectral prominences while [e] has a single narrow low-frequency spectral peak.

The present study asks if formant averaging in vowel perception generalizes to adjacent oral and nasal vowel formants. Our purpose was to determine whether the perception of height in nasal vowels involves spectral integration of the first oral formant, F1, and the first nasal formant, FN.

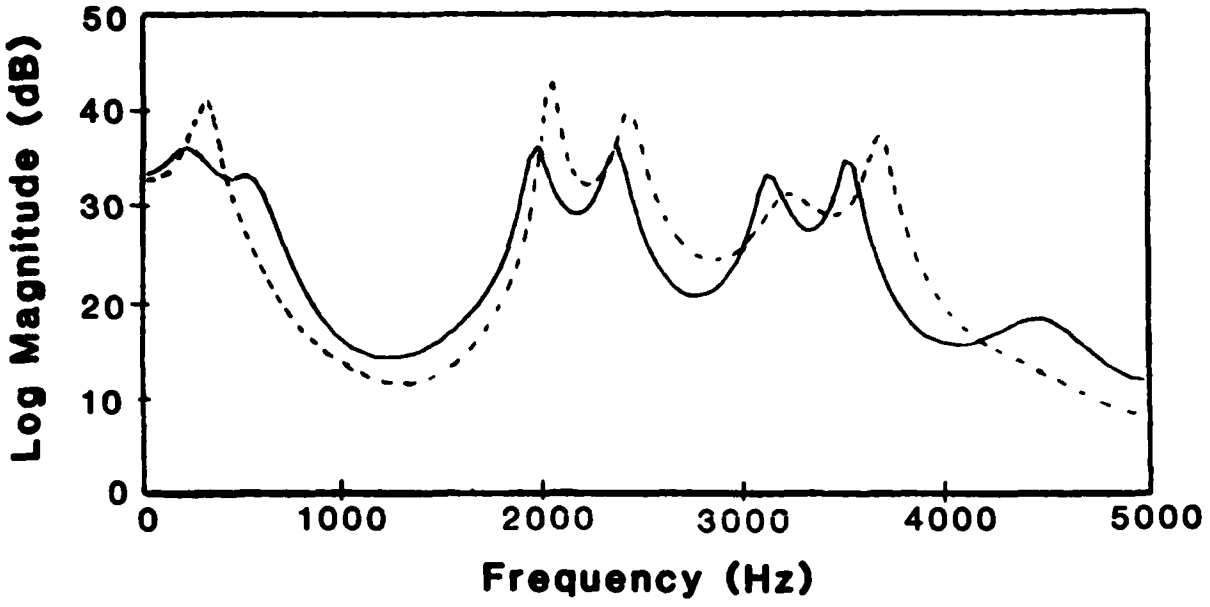


Figure 1. LPC spectra of nasal [ẽ] (solid curve) and oral [e] (dashed curve) produced by a Hindi speaker. The nasal vowel spectrum has two broad spectral prominences in the low-frequency region while the oral vowel spectrum has a single narrow low-frequency spectral peak.

The oral vowel studies reviewed above might lead us to expect F1-FN averaging since the distance between F1 and FN in many nasal vowels is less than 3.5 Bark, i.e., less than the critical distance found for spectral integration of oral vowel components. (For example, the distance between the first two spectral peaks of nasal [ẽ] in Figure 1 is roughly 2.8 Bark.) Previous nasal vowel research also points toward possible F1-FN integration. Joos (1948) suggested that French /ɛ̃/ sounded like [ã] because the average frequency of F1 and FN in nasal /ɛ̃/ corresponds to F1 in oral /æ/. Similarly, Fant (1960) and Wright (1980) speculated that shifts in perceived vowel height accompanying nasal coupling might be due to the additional low-frequency nasal resonance.

Method

Stimulus Materials

The stimulus materials were five sets of nasal and oral vowels generated on the Haskins serial software formant synthesizer. Each 360-ms stimulus consisted of steady-state vowel formants, with fundamental frequency and amplitude decreasing over the final 120 ms.

The five nasal vowel stimuli, [ĩ ẽ æ̃ ã õ], were synthesized by adding a pole-zero pair in the vicinity of the first pole to the five-pole transfer function for an oral vowel. The spectral characteristics of the synthetic nasal vowels were based on FFT and LPC analyses of natural vowel tokens from several languages (Beddor, 1983). Autoregressive LPC spectra of the synthesized nasal vowels are shown in Figure 2, along with the measured frequencies of the first two spectral peaks. The labels assigned to these peaks are to be interpreted with caution, since identifying the "first oral formant" and the

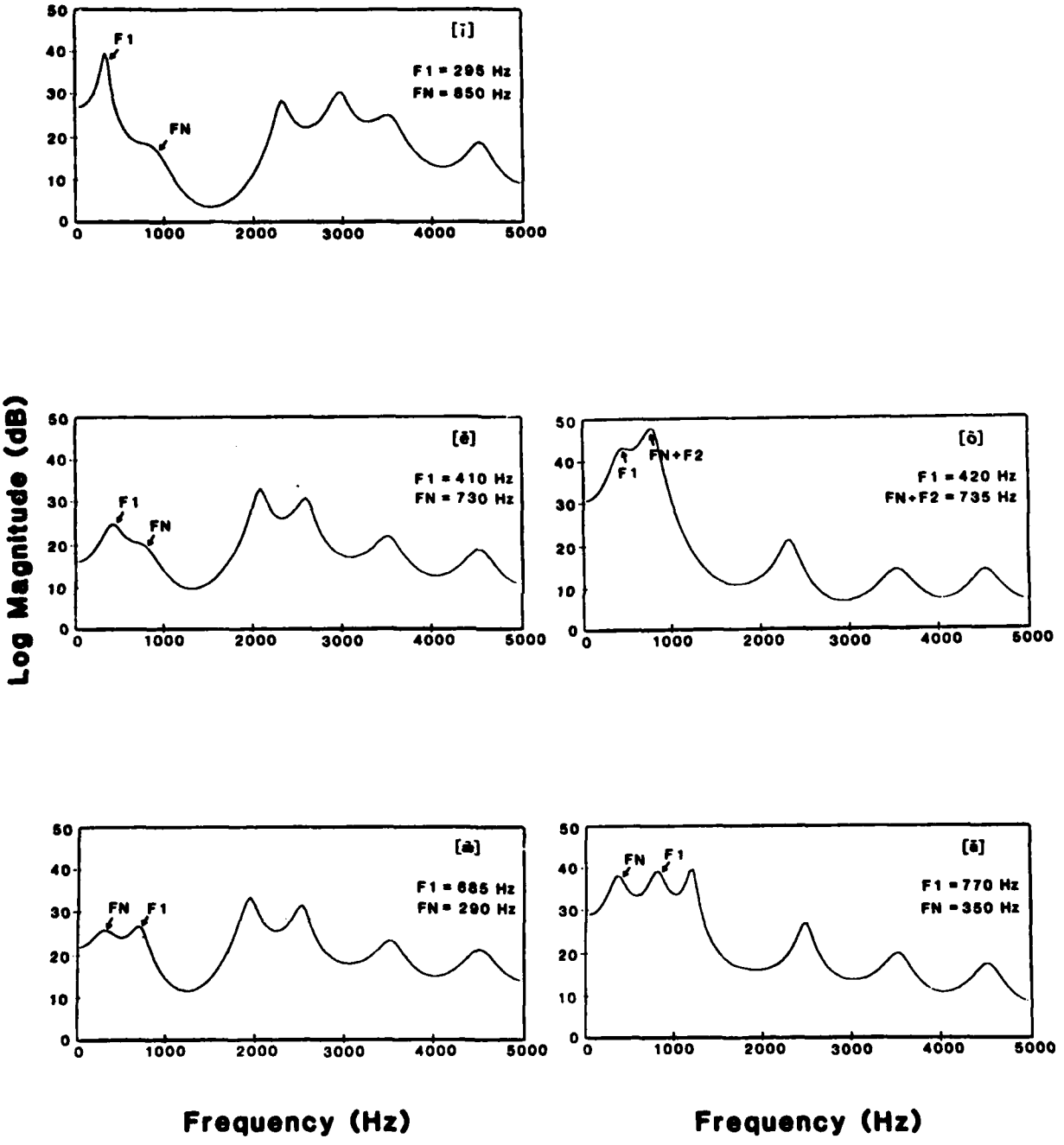


Figure 2. Spectra of the five synthetic nasal vowel stimuli.

extra "nasal formant" of nasal vowels is a terminological problem (see Stevens, Fant, & Hawkins, in press). The convention adopted here is to label as "F1" the first peak in the high and mid nasal vowels, [ĩ], [ẽ], and [õ], and the second peak in the low nasal vowels, [æ̃] and [ã] (these "F1" values being close to typical F1 frequencies for the oral vowels [i], [e], [o], [æ], and [a]). That is, F1 frequency was less than FN frequency in high and mid vowels and greater than FN frequency in low vowels, which is consistent with the acoustic theory of vowel nasalization (Fant, 1960; Fujimura and Lindqvist, 1971) as well as previous analyses of natural nasal vowel tokens (e.g., Fujimura, 1961; Wright, 1980). The added zero was set between the first oral pole and the additional pole for all nasal vowels except high [ĩ], where the zero separated the additional pole and the second oral pole (see Fujimura, 1961; Maeda, 1982).

For each nasal vowel, a continuum of oral vowels was constructed by omitting the extra pole-zero pair. Within each oral continuum, stimuli were identical to each other except for the frequency of F1, which was systematically varied as shown in Table 1. F1 step-size in each continuum was approximately 10% of the average F1 frequency for that vowel set. (Thus step sizes were larger for lower vowels, e.g., F1 step-size was 32 Hz for i, 45 Hz for e, and 60 Hz for æ.) The F1 range of each oral continuum included two vowels of special interest. One of these oral vowels was an "F1 match": the frequency of its first formant was the same as the F1 frequency of the corresponding nasal vowel. (This can be seen by comparing the oral vowel F1 values designated by * in Table 1 with the nasal vowel F1 values in Figure 2.[2]) A second oral vowel from each of the five series was a "centroid match" (** in Table 1);[3] this stimulus matched the corresponding vowel on a specific measure of center of gravity.

The centroid of a vowel is a measure of the center of gravity calculated from the LPC spectrum of that vowel. The centroid (CEN) function computes the mean frequency of the area under the spectral curve within specified frequency and magnitude ranges according to the formula

$$X_{CEN} = \frac{\sum_{i=1}^{n} (X_i Y_i)}{\sum_{i=1}^{n} (Y_i)}$$

where X = frequency (Hz) and Y = log magnitude (dB). Figure 3 demonstrates the operation of the centroid function for nasal [ẽ]. The left and right vertical bars delimit the frequency range of 100-1100 Hz and the connecting horizontal bar sets the lower magnitude limit. The spectral curve forms the upper magnitude limit. The center frequency or centroid of this area, 526 Hz, is shown by the dashed vertical line. The frequency and magnitude ranges selected in this study were based on analyses of over 800 natural-speech tokens of oral and nasal vowels (see Beddor, 1983, for discussion of these ranges). The frequency range of 100-1100 Hz was used for all vowel stimuli except for the low central vowels, for which the upper limit was extended to 1400 Hz.[4] The lower magnitude limit was determined separately for each stimulus and was set just below the lowest point in the 100-1100 (or 1400) Hz portion of the spectral curve. The area measured by the centroid function included F1 in all vowels, but also FN in the nasal vowels and F2 in the non-front (oral and nasal) vowels.

Table 1

F1 values (in Hz) for the oral vowel sets.

	Stimulus Number									
	1	2	3	4	5	6	7	8	9	10
i	263	295*	327	359	391**	423				
e	275	320	365	410*	455	500	545**	590	635	680
æ	390	450	510**	570	630	690*	750	810		
a	420	490	560**	630	700	770*	840	910		
o	300	340	380	420*	460	500**	540	580		

* =F1 match ** =Centroid match

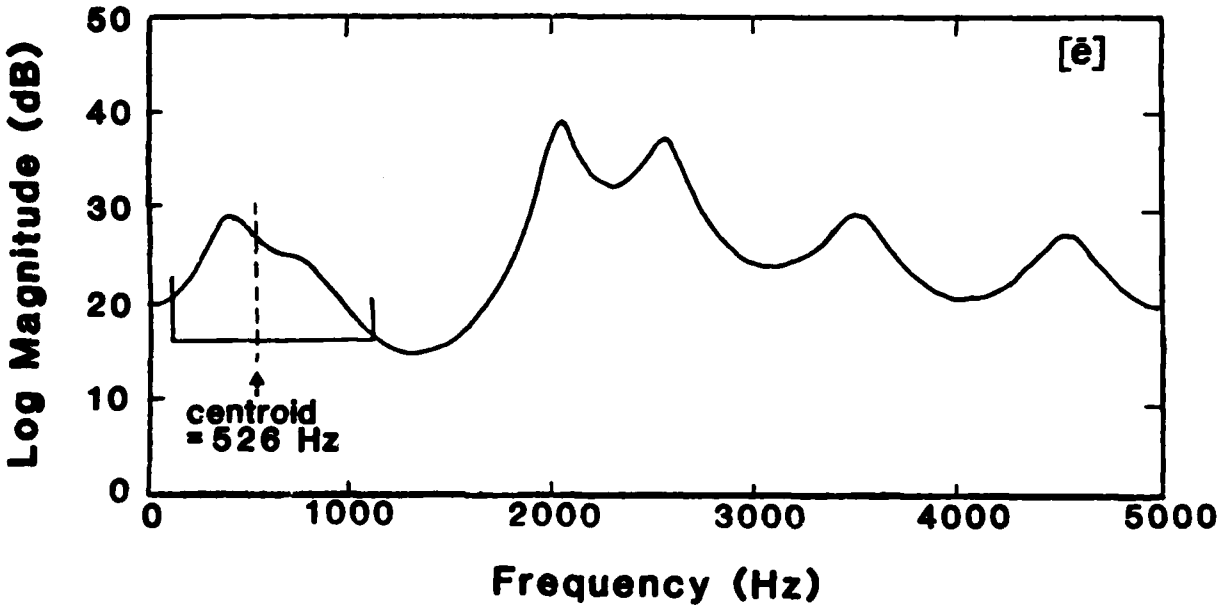


Figure 3. Illustration of the centroid function using the mid front nasal vowel stimulus, [ẽ]. The figure indicates the region of the spectrum analyzed by the centroid function: the vertical bars delimit the 100-1100 Hz frequency range, the horizontal bar sets the lower magnitude limit, and the spectral curve forms the upper magnitude limit. The dashed line marks the center frequency or centroid of this region.

Figure 4 compares the stimuli designated "F1 match" and "centroid match" from vowel set æ. In the upper panel, the F1 match, we see that the frequency of the first peak in the oral vowel spectrum (dashed curve) and the frequency of the second peak in the nasal vowel spectrum (solid curve) are the same. In contrast, in the centroid match in the lower panel, the first peak in the oral vowel straddles the two low-frequency peaks of the nasal vowel; while these two spectra share no peak frequency in the first region of spectral prominence, the center frequency of this region is the same in the two spectra.

Figure 4 also shows that, in vowel set æ, the oral vowel of the centroid-matched pair has a lower F1 frequency than the oral vowel of the F1-matched pair. This is also true of the low vowel set a, as indicated by the values in Table 1. In contrast, in the non-low vowel sets, i, e, and o, the centroid match has a higher F1 frequency than the F1 match. This is due to the location of the first nasal formant relative to the first oral formant in the nasal vowels (see Figure 2): when FN frequency is less than F1 frequency, as in low nasal vowels, FN pulls down the center of gravity; when FN is greater than F1, as in high and mid nasal vowels, FN pulls up the center of gravity.

Subjects

Twenty paid student volunteers participated in the experiment. All were native speakers of American English with no known hearing loss and no expertise in phonetics. Although several of the subjects had studied a language in which the oral-nasal contrast in vowels is distinctive (e.g., French, Polish), this background had no apparent effect on their results.

Procedure

Test sequences for the five vowel sets consisted of pairs of oral and corresponding nasal vowels. For each set, two types of ordered sequences were made: ascending sequences (i.e., each oral stimulus from 1 through n paired with the nasal standard) and descending sequences (i.e., oral-nasal pairs from n through 1). A pilot study in which listeners selected the "best-match" oral-nasal pair from these sequences showed that matches tended to fall in the middle of the vowel set. To eliminate clustering of responses in the center of each vowel set, three truncated ordered sequences for each vowel set were constructed from the full ascending and descending sequences. The truncated sequences contained the following oral stimuli (paired with the corresponding nasal vowel): i: 1-5 (twice), 2-6; e: 1-8, 2-9, 3-10; æ: 1-7 (twice), 2-8; a: 1-6, 2-7, 3-8; o: 1-6, 2-7, 3-8. The three truncated versions of each of the five vowel sets were arranged in random order, for a total of 15 trials. The inter-stimulus interval between members of an oral-nasal pair was .5 s and the interval between pairs in the ordered sequences was 1 s; subjects controlled intervals across sequences and trials.

Before testing, subjects were given a brief description of the kinds of vowel stimuli to be presented. Subjects were told that they would hear 15 sets of vowels, each set consisting of several vowel pairs. They were informed that the first member of each pair varied across the series while the second member stayed the same and that these pair members were "oral vowels" and "nasal vowels," respectively. It was explained that nasal vowels usually

occur in English in the context of m or n, e.g., mom (versus the oral vowel in Bob), man (versus bad), and moan (versus boat).

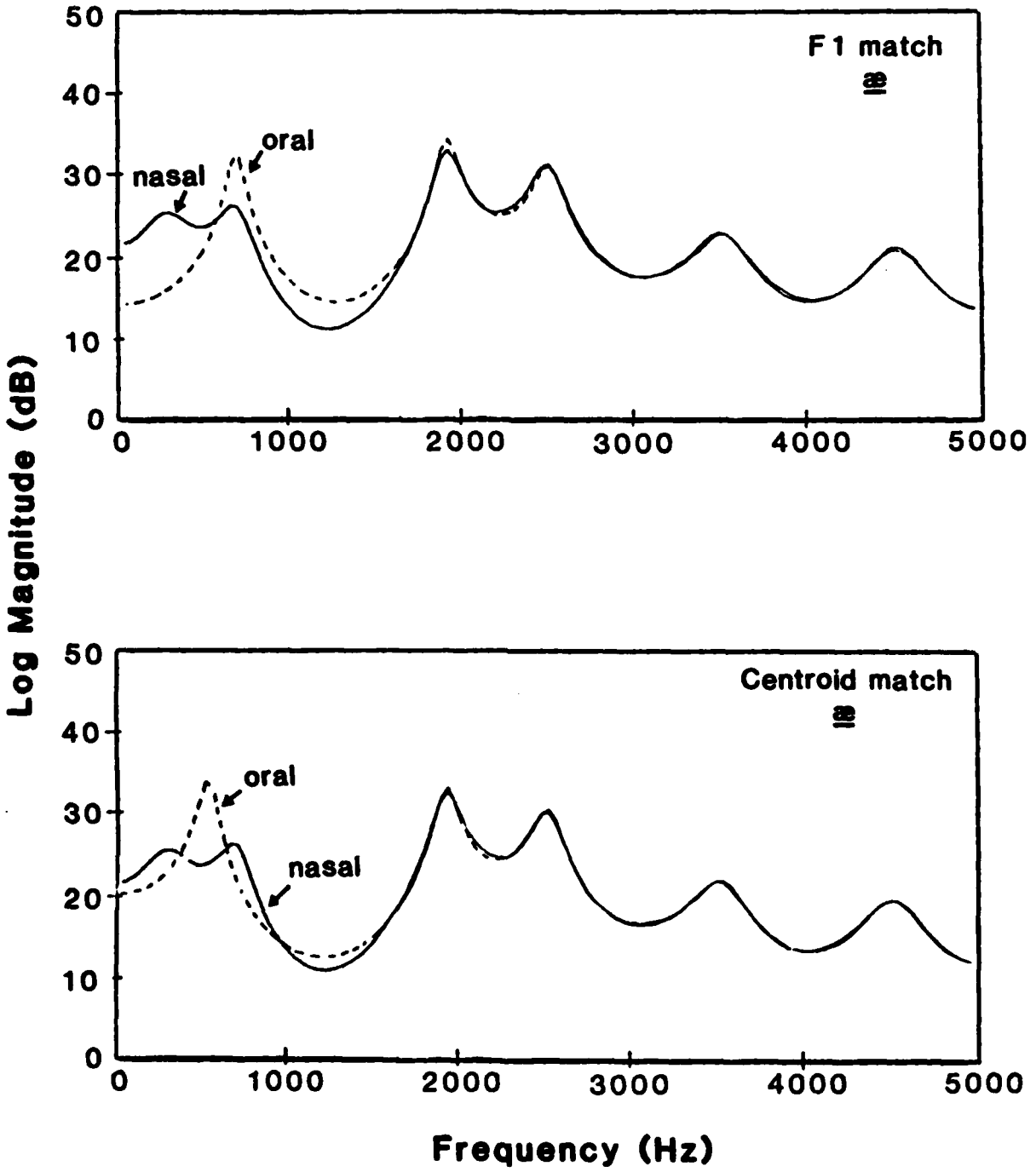


Figure 4. Spectra of the F1-matched stimulus pair (upper panel) and the centroid-matched stimulus pair (lower panel) for the low front vowel set æ.

Subjects were tested individually in a sound-attenuated booth. Stimuli were presented binaurally over TDH-39 earphones with an interactive computer program. At the onset of each of the 15 trials, the program presented the ascending and descending truncated sequences for that vowel set. (The relative order of ascending and descending sequences was counter-balanced across trials.) The subject could then request repetitions of either of the sequences or of individual oral-nasal pairs from the sequence. For each trial, a subject was instructed to select that pair in which the oral vowel was the most similar to the nasal standard; this "best-match" pair was circled on a printed score sheet. A subject was encouraged to listen to the sequences and to individual pairs as many times as needed to feel confident about the best-match decision. Average testing time was approximately 45 minutes.

Results

The histograms in Figure 5 show subjects' responses to the five vowel sets, i, e, æ, a, and o. As there was no apparent effect of truncation, responses to the three truncated versions of each vowel set were pooled. The data therefore represent 60 responses (20 subjects X 3 truncations) per vowel set. Oral vowel stimulus number is on the ordinate and percent best-match responses on the abscissa. The F1 match in each vowel set is indicated by * and the centroid match by **.

Figure 5 shows that subjects' best-match responses to each vowel set are spread over several stimulus pairs. Of special interest here are the F1- and centroid-matched pairs. It was hypothesized that if perceived nasal vowel height were determined by center of gravity, then the perceptually most similar oral-nasal pair in each vowel set would be the centroid-matched pair. If, however, perceptual integration of F1 and FN did not occur, then the most similar pair might be expected to be the F1-matched pair.

As seen in Figure 5, the F1-matched oral-nasal pair in vowel set i accounted for over 70% of subjects' responses. But in the remaining four vowel sets, subjects perceived the F1-matched pair as the most similar pair only 2% to 12% of the time. For each of the five vowel sets, a t-test of the difference between the stimulus number of the F1 match and the mean stimulus value of each subject's responses showed that responses differed significantly from the F1-matched vowel pair, i, $t(19) = 2.68$, $p < .05$; e, $t(19) = 11.87$, $p < .01$; æ, $t(19) = 15.88$, $p < .01$; a, $t(19) = 14.45$, $p < .01$, and o, $t(19) = 10.97$, $p < .01$. These findings are consistent with the data of Wright (1980), which showed that perceptual effects of nasalization on vowel height were not always a function of acoustic effects of nasalization on first formant frequency.

Although listeners generally did not match oral and nasal vowels on the basis of first formant frequency, they also tended not to choose the centroid-matched pairs as perceptually similar. In the mid and low vowel sets, the most frequently-chosen oral-nasal pair fell between the F1 and centroid pairs. This modal best-match response was closer to the centroid for mid front e, but closer to F1 for the low vowels æ and a. However, due to the centroid skew of the æ, a, and o distributions, subjects' mean response (given in Figure 5) was closer to the centroid than to F1 for all four non-high vowel sets. A t-test for each vowel set compared the difference between the stimulus number of the centroid match and each subject's mean response to the difference between the F1 match and mean responses. The analyses showed that

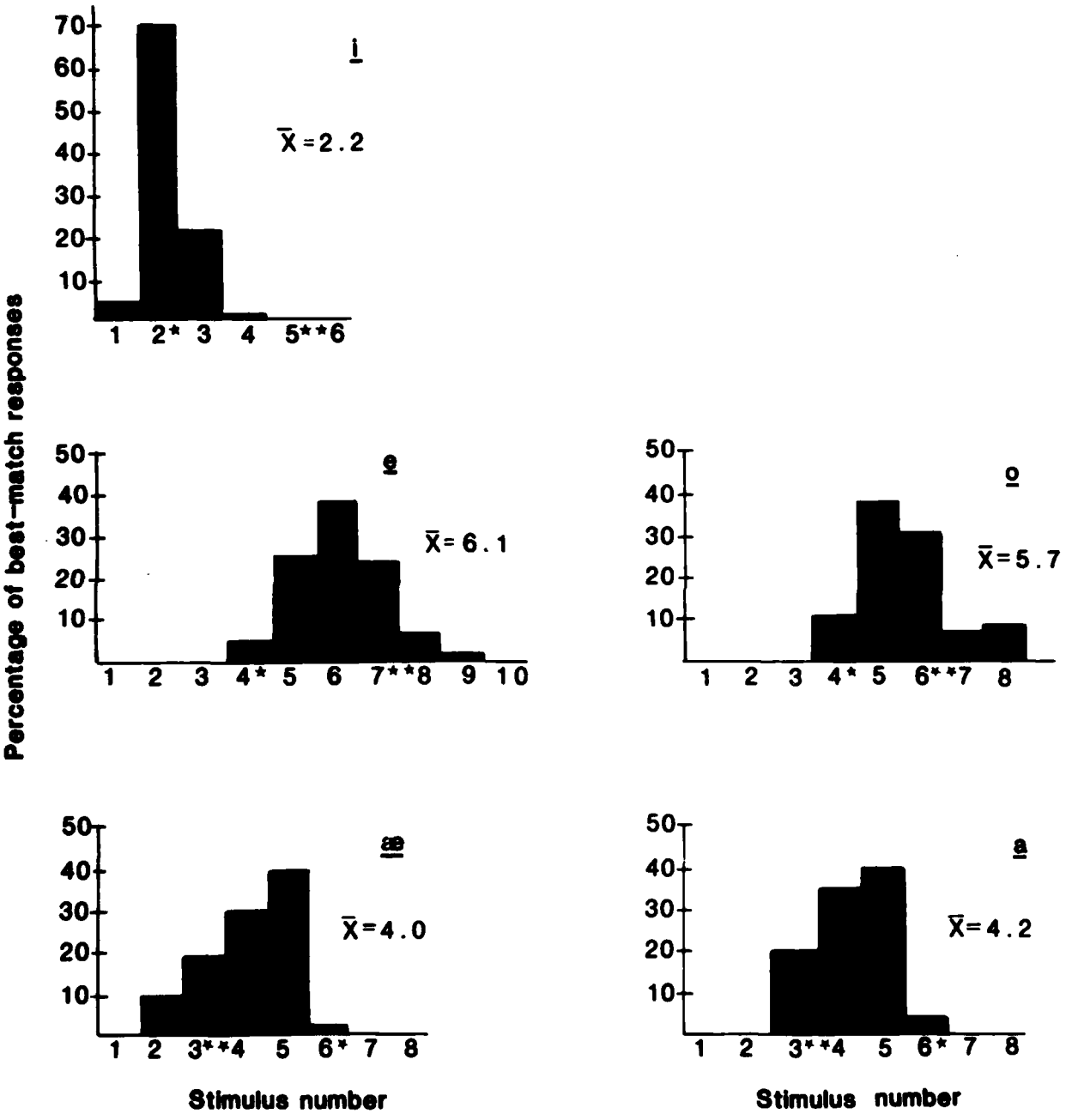


Figure 5. Percent "best-match" responses to the oral-nasal vowel pairs for the vowel sets i, e, æ, a, and o, where * = F1-matched pair and ** = centroid-matched pair.

perceptually-similar pairs of oral and nasal vowels were significantly closer to the centroid-matched pair than to the F1-matched pair in the four non-high vowel series, e, æ, o, t(19) = 3.27, 3.41, 4.87, respectively, $p < .01$; and a, $t(19) = 2.24$, $p < .05$. Only in the high vowel set i were listeners' responses significantly closer to the vowel pair matched for F1 frequency, $t(19) = 18.15$, $p < .01$.

Discussion

The purpose of this experiment was to determine whether spectral integration of the first oral and nasal formants occurs in the perception of nasal vowels such that perceived nasal vowel height correlates with the center of the first region of spectral prominence rather than with the frequency of the first formant. The method used to elicit height judgments from phonetically-naive subjects required listeners to select from a continuum of oral vowels the vowel that was perceptually most similar to a nasal vowel standard. Since the oral stimuli differed from the nasal standard only in low-frequency spectral characteristics, the selected oral match was taken as an indication of the perceived height of the nasal vowel. The results suggest that perception of nasal vowel height, as measured by this paradigm, involves integration of low-frequency spectral prominences. Perceived nasal vowel height was not determined solely by the first formant: with the exception of high [ĩ], F1 accounted for very few of the listeners' responses. Rather, listeners' responses showed very consistent deviations from F1: when the frequency of FN was less than the frequency of F1, as in low [æ̃] and [ã], the closest oral match had a relatively low F1 frequency; when FN frequency was greater than F1 frequency, as in mid [ẽ] and [õ], the selected oral match had a relatively high F1. In all four of these vowel sets, the selected F1 frequency of the oral vowel was intermediate relative to the F1 and FN frequencies of the nasal vowel. Even for high [ĩ], over 80% of the non-F1 responses were pulled in the direction of the nasal formant. Thus our data provide empirical support for previous speculations that the relative positions of the first oral and nasal formants might influence perceived nasal vowel height (Fant, 1960; Joos, 1948; Wright, 1980).

The finding that perceived nasal vowel height was not determined by the frequency of a single low-frequency spectral peak but rather involved apparent integration of low-frequency spectral components demonstrates the center-of-gravity effect in the perception of nasal vowel height. The high front nasal vowel, however, did not show strong evidence of perceptual integration of F1 and FN: the majority of listeners' responses to [ĩ] points toward F1 frequency as determining perceived height. A possible explanation for this difference between the high and non-high vowels lies in the distance between F1 and FN frequencies in [ĩ] versus [ẽ], [æ̃], [ã], and [õ]. As noted above, Chistovich and Lublinskaya (1979) and Syrdal and Gopal (1983) report that formant integration does not occur in oral vowels (i.e., the center-of-gravity effect disappears) when formant distance exceeds 3.5 Bark. In our stimuli, the separation between F1 and FN in the mid and low nasal vowels was 2.5 and 3.4 Bark, respectively, while the separation for the high nasal vowel was 4.5 Bark. F1-FN integration in the mid and low, but not the high, nasal vowels is therefore consistent with previous oral vowel findings.

While the center-of-gravity effect is apparent in the mid and low vowel data, it is also clear that perceived nasal vowel height did not correspond exactly with our measure of center of gravity, the centroid. Although obtained matches between oral and nasal vowels were significantly closer to the centroid-matched pairs than to the F1-matched pairs, 38% to 75% of the responses to the non-high vowel sets fell between the F1- and centroid-matched pairs. This bias towards F1 in listeners' judgments indicates that, for the centroid to reflect perceived vowel height, F1 should be given more weight than in the current measure.[5] Note, however, that such a revision is not simply a matter of increasing the weight of the lowest-frequency spectral peak, since in low [æ̃] and [ã], F1 was the second, rather than the first, spectral prominence. Although identification of the spectral prominence corresponding to F1 is problematic in nasal vowels, this problem does not change our finding that subjects' responses were higher than the centroid for low nasal vowels but lower than the centroid for non-low nasal vowels. Furthermore, the F1 bias cannot be accounted for by increasing the weight of the higher-magnitude spectral peak, since the magnitude of the second peak was greater than the magnitude of F1 in mid [õ]. It appears, then, that no simple weighting of spectral components in terms of their frequency and magnitude will account for perceived center of gravity in nasal vowels. Whether oral vowels show a similar discrepancy between perceived center of gravity and the centroid is currently under investigation.

In summary, although our measure of center of gravity needs to be revised, the results clearly evidence the center-of-gravity effect in the perception of nasal vowel height. Previous studies with oral vowels have shown that vowel formants are integrated over frequency intervals which are broader than a critical band (Bladon, 1983; Chistovich & Lublinskaya, 1979; Syrdal & Gopal, 1983). Our findings with the first oral and nasal formants of nasal vowels show that nasal vowel formant energy is also integrated over relatively wide frequency intervals. Whether the critical distance for formant averaging is the same in nasal vowels as in oral vowels needs further study. The data presented here, however, are consistent with the critical distance of 3.5 Bark previously reported for oral vowels.

References

Beddor, P. (1983). Phonological and phonetic effects of nasalization on vowel height. Bloomington: Indiana University Linguistics Club.

Bedrov, Ya., Chistovich, L., & Sheikin, R. (1978). Frequency position of the "center of gravity" of formants as a useful feature in vowel perception. Soviet Physics Acoustics, 24, 275-278.

Bladon, A. (1983). Two-formant models of vowel perception: Shortcomings and enhancements. Speech Communication, 2, 305-313.

Bladon, R. A., & Fant, G. (1978). A two-formant model and the cardinal vowels. Speech Transmission Laboratory Quarterly Progress and Status Report (Stockholm Royal Institute of Technology), 1, 1-8.

Carlson, R., Fant, G., & Granström, B. (1975). Two-formant models, pitch, and vowel perception. In G. Fant & M. Tatham (Eds.), Auditory analysis and perception of speech (pp. 55-82). New York: Academic Press.

Carlson, R., Granström, B., & Fant, G. (1970). Some studies concerning perception of isolated vowels. Speech Transmission Laboratory Quarterly Progress and Status Report (Stockholm Royal Institute of Technology), 2-3, 19-35.

Chistovich, L., & Lublinskaya, V. (1979). The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli. Hearing Research, 1, 185-195.

Chistovich, L., Sheikin, R., & Lublinskaya, V. (1979). 'Centers of gravity' and spectral peaks as the determinants of vowel quality. In B. Lindblom & S. Ohman (Eds.), Frontiers of speech communication research (pp. 143-157). New York: Academic Press.

Delattre, P. (1954). Les attributs acoustiques de la nasalité vocalique et consonantique. Studia Linguistica, 8, 103-108.

Delattre, P., Liberman, A., Cooper, F., & Gerstman, L. (1952). An experimental study of the acoustic determinants of vowel color; Observations on one- and two-formant vowels synthesized from spectrographic patterns. Word, 8, 195-210.

Fant, G. (1960). Acoustic theory of speech production. The Hague: Mouton.

Fujimura, O. (1961). Analysis of nasalized vowels. Quarterly Progress Report (MIT Research Laboratory of Electronics), 62, 191-192.

Fujimura, O., & Lindqvist, J. (1971). Sweep-tone measurements of vocal-tract characteristics. Journal of the Acoustical Society of America, 49, 541-558.

Fujisaki, H., & Kawashima, T. (1968). The roles of pitch and higher formants in the perception of vowels. IEEE Transactions on Audio and Electroacoustics, AU-16, 73-77.

Hamada, M. 1983. Nasal vowel identification using LPC-based formant analysis. Speech Communication Group Working Papers (MIT Research Laboratory of Electronics), 3, 41-54.

House, A., & Stevens, K. (1956). Analog studies of the nasalization of vowels. Journal of Speech and Hearing Disorders, 21, 218-232.

Joos, M. (1948). Acoustic phonetics (Language Monograph 23). Baltimore: Linguistic Society of America at Waverly Press.

Ladefoged, P. (1982). A course in phonetics (2nd ed.). Chicago: Harcourt Brace Jovanovich.

Maeda, S. (1982). Acoustic cues for vowel nasalization: A simulation study. Journal of the Acoustical Society of merica, 72, S102. (Abstract)

Miller, R. (1953). Auditory tests with synthetic vowels. Journal of the Acoustical Society of America, 25, 114-121.

Mrayati, M. (1975). Etude des voyelles françaises. Bulletin de l'Institut de Phonétique de Grenoble, 4, 1-26.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 24, 175-184.

Schroeder, M., Atal, B., & Hall, J. (1979). Objective measure of certain speech signal degradations based on masking properties of human auditory perception. In B. Lindblom & S. Ohman (Eds.), Frontiers of speech communication research (pp. 217-229). New York: Academic Press.

Stevens, K., Fant, G., & Hawkins, S. (in press). Some acoustical and perceptual correlates of nasal vowels. In R. Channon & L. Shockey (Eds.), Festschrift for Ilse Lehiste.

Syrdal, A., & Gopal, H. (1983). Perceived critical distances between F1-F0, F2-F1, F3-F2. Journal of the Acoustical Society of America, 74, S88-S89. (Abstract)

Traunmüller, H. (1981). Perceptual dimension of openness in vowels. Journal of the Acoustical Society of America, 69, 1465-1475.

Wright, J. (1980). The behavior of nasalized vowels in the perceptual vowel space. Report of the Phonology Laboratory (University of California, Berkeley), 5, 127-163.

Footnotes

[1]The Bark scale divides the audible frequency range into units of critical bands, where 1 Bark equals one critical band. The relationship of Hertz to Bark is expressed in the following equation from Schroeder, Atal, and Hall (1979)

$$f = 650 \sinh(x/7)$$

where f is frequency in Hz and x is frequency in Bark.

[2]Since there is a problem in identifying the first oral versus the first nasal formant of nasal vowels, we might ask whether the F1 match indeed matches first oral formant frequencies of the oral and nasal vowels or whether it might be a F1-FN match in some vowel sets. One way to avoid this issue would be to extend the F1 range covered by each oral vowel continuum to include the frequencies of both F1 and FN of the corresponding nasal vowel. However, a pilot study with such extended continua indicated that pairs in which F1 frequency of the oral vowel matched what we have labeled "FN" frequency of the nasal vowel were very poor perceptual matches. Inasmuch as these extended series were unnecessarily long, the "FN" end of the series was omitted in the actual experiment.

[3]For all vowel sets, matches between oral and nasal vowels in F1 and centroid values were based on measurements of LPC spectra of these vowels. In the LPC analysis, 14 predictor coefficients were calculated for each oral vowel and 18 for each nasal vowel. To verify the LPC measures, F1 and centroid values were also obtained from FFT spectra of the vowel tokens. These frequencies were within 15 Hz of the LPC measures.

[4]For a single frequency range to be applied in each vowel set, a rather broad frequency range was necessitated by the variation in F1 frequency in the oral continua (the F1 frequencies of the endpoint stimuli in a continuum were up to 490 Hz apart). This broad range, however, is not meant to imply that perceivers average spectral information over a 1000 Hz range. A more accurate interpretation is that these frequencies might be relevant to perception of vowel height; additional research is of course necessary to determine the limits of the relevant frequency range.

[5]Similarly, Carlson, Fant, and Granström (1975) reported that efforts to calculate F2' as a linearly weighted mean frequency of F2, F3, and F4 were unsuccessful. Their revised formula gave greater weight to F2 when F2 was close to F3 but greater weight to F3 and F4 when F2 and F3 were far apart.

RELATIVE POWER OF CUES: FO SHIFT VS. VOICE TIMING*

Arthur S. Abramson† and Leigh Lisker††

Background

The acoustic features that bear information on the identity of phonetic segments are commonly called cues to speech perception. These cues do not typically have one-to-one relationships with phonetic distinctions. Indeed, research usually shows more than one cue to be pertinent to a distinction, although all such cues may not be equally important. Thus, if two cues, x and y, are relevant for a distinction, it may turn out that for any value x, a variation of y will effect a significant shift in listeners' phonetic judgments, but that there will be some values of y for which varying x will have negligible effect on phonetic judgments. We say then that y is the more powerful cue.

A good deal of evidence now exists to show that the timing of the valvular action of the larynx relative to supraglottal articulation is widely used in languages to distinguish homorganic consonants. The detailed properties of the distinctions thus produced depend on glottal shape and concomitant laryngeal impedance or stoppage of airflow, as well as on the phonatory state of the vocal folds. Such acoustic consequences as the presence or absence of audible glottal pulsing during consonant closures or constrictions, the turbulence called aspiration between consonant release and onset or resumption of pulsing, and damping of energy in the region of the first formant, have all been subsumed by us (Lisker & Abramson, 1964, 1971) under a general mechanism of voice timing. In utterance-initial position, the phonetic environment in which consonantal distinctions based on differences in the relative timing of laryngeal and supraglottal action have been most often studied, this phonetic dimension has commonly been referred to as voice onset time or VOT.

Although the acoustic features just mentioned, and perhaps some others, may be said to vary under the control of the single "mechanism" of voice timing, it is of course possible, by means of speech synthesis, to vary them one at a time to learn which of them are perceptually more important. We must not forget, however, that such experimentation involves pitting against one another acoustic features that are not independently controlled by the human speaker.

---

*Also to appear in V. Fromkin (Ed.), Phonetic linguistics. New York: Academic Press.

†Also University of Connecticut.

††Also University of Pennsylvania.

Acknowledgment. This work was supported by Grant HD-01994 from the National Institute of Child Health and Human Development to Haskins Laboratories. An oral version was presented at the Tenth International Congress of Phonetic Sciences, Utrecht, 1-6 August, 1983.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

A relevant feature not so far mentioned is the fundamental frequency (F0) of the voice. If we assume a certain F0 contour as shaped by the intonation or tone of the moment, there is a good correlation between the voicing state of an initial consonant and the F0 height and movement at the beginning of that contour (House & Fairbanks, 1953; but see also O'Shaughnessy, 1979, for complications). After a voiced stop, F0 is likely to be lower and shift upward, while after a voiceless stop it will be higher and shift downward (Lehiste & Peterson, 1961). Although the phenomenon has not been fully explained, it is at least apparent that it is a function of physiological and aerodynamic factors associated with the voicing difference.

The data derived from the acoustic analysis of natural speech can be matched by experiments with synthetic speech that demonstrate that F0 shifts can influence listeners' judgments of consonant voicing (Fujimura, 1971; Haggard, Ambler, & Callow, 1970; Haggard, Summerfield, & Roberts, 1981). Of further interest in this connection is the claim that phonemic tones have developed in certain language families through increased awareness of these voicing-induced F0 shifts and their consequent promotion to distinctive pitch features under independent control in production (Hombert, Ohala, & Ewan, 1979; Maspero, 1911).

Our motivation for the present study was to put F0 into proper perspective as one of a set of potential cues to consonant voicing coordinated by laryngeal timing. After all, our own earlier synthesis (Abramson & Lisker, 1965; Lisker & Abramson, 1970) yielded quite satisfactory voicing distinctions without F0 as a variable. In addition, Haggard et al. (1970) may have exaggerated its importance in the perception of natural speech by their use of a frequency range of 163 Hz, one very much greater than, for example, the range of less than 40 Hz found for English stop productions by Hombert (1975). We set out to test the hypothesis that the separate perceptual effect of F0 is small and dependent upon voice timing, while the dependence of the voice timing effect on F0 is virtually nil. We used native speakers of English as test subjects.

Procedure

Making use of the Haskins Laboratories formant synthesizer, we prepared a pattern appropriate to an initial labial stop followed by a vowel [ɑ]. Variants of this pattern were then synthesized with VOT values of 5, 20, 35, and 50 ms after the simulated stop release. These values were chosen because of earlier work (Figure 1) that determined English voicing judgments for a VOT continuum ranging from 150 ms before release to 150 ms after release. This range of VOT values was sampled at 10 ms intervals, except for the span from 10 ms before release to 50 ms after release, which was sampled at 5 ms intervals. Those stimuli for which voice onset followed release, i.e., to the right of 0 ms on the abscissa, had noise-excited upper formants during the interval between the burst at VOT = 0 and the onset of voice. In the labial data at the top of the figure the perceptual crossover point between /b/ and /p/ falls just after 20 ms of voicing lag. Thus we expected that the extreme values of our more limited range would be heard as unambiguous /b/ and /p/, given an unchanging F0, while the category boundary, lying somewhere between, might be shifted one way or the other as the F0 was varied. In addition to a set of VOT variants having an F0 fixed at 114 Hz, we imposed onset frequencies of 98, 108, 120, and 130 Hz, values commensurate with ranges reported for natural speech (Hombert, 1975; House & Fairbanks, 1953; Lea, 1973; Lehiste & Peter-

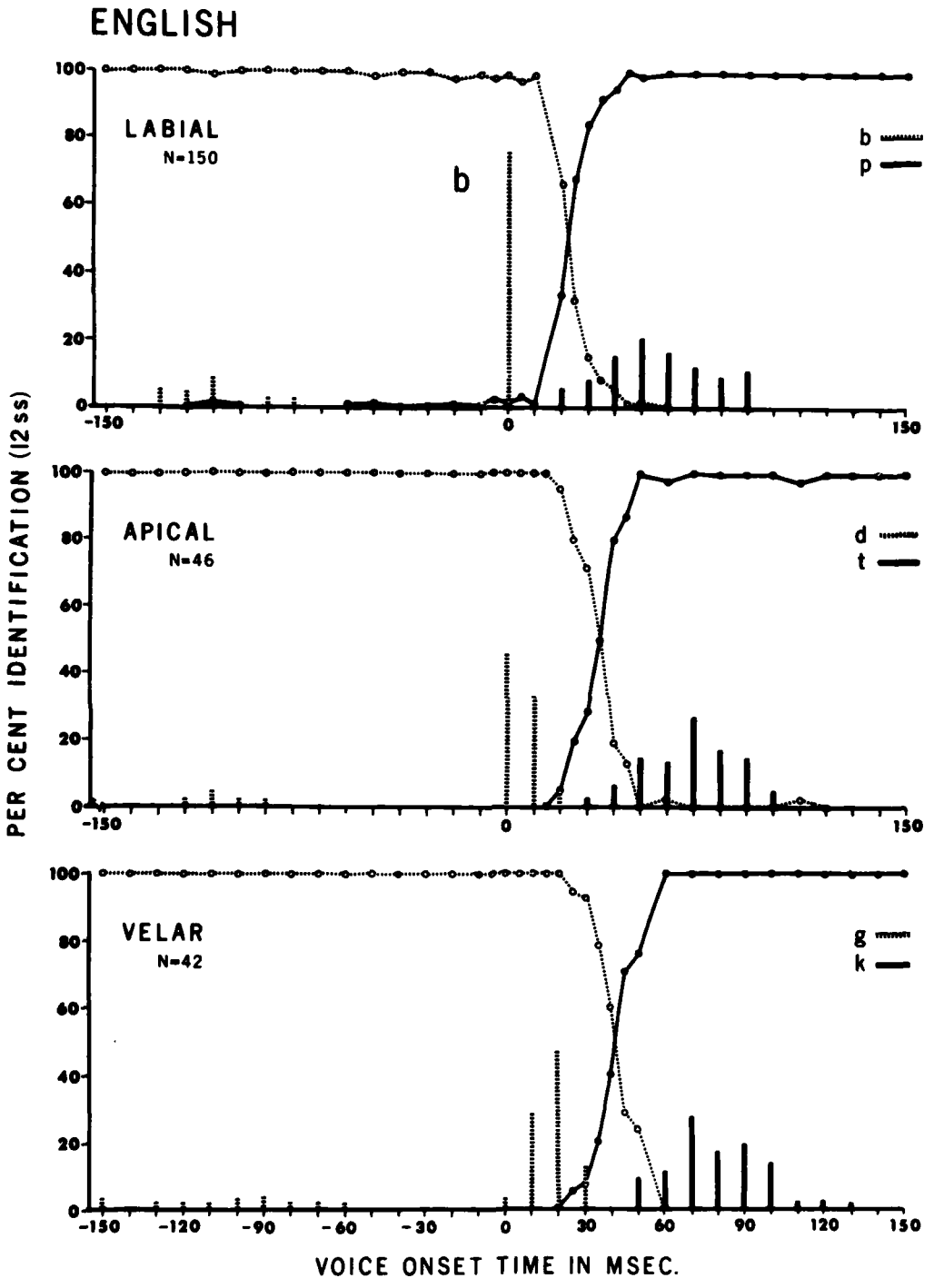


Figure 1. English voicing judgments for stops varying in VOT. Below each pair of curves, a histogram (from Lisker & Abramson, 1964) of frequency distributions of VOT in speech. (Reproduced from Lisker & Abramson, 1970.)

son, 1961). That is, the F0 at voicing onset for each variant began at one of those frequencies and shifted upward or downward to a level of 114 Hz where it stayed for the rest of the syllable. These F0 shifts were of three durations, 50, 100, and 150 ms. These fit with our own cursory observations and bracket the value of 100 ms found by Hombert (1975). We recorded the resulting 52 stimuli--two tokens of each--in three randomizations and played the tapes to 11 native speakers of English for labeling as /b/ or /p/. The subjects, three women and eight men, represented a wide variety of regional dialects, ten in the United States and one in Britain.

Results

The overall results are shown in Figure 2. The three panels are for the durations of F0 shift. The abscissa of each panel shows the four VOT values, while the ordinate gives the percentage identified as /p/ for each VOT. The coded line standing for the variants with a flat F0 of 114 Hz is, of course, a plot of the same data in all three panels. The 50% perceptual crossover point for the flat F0 falls at about 25 ms of VOT. This is consistent with the results for the more finely graded series of stimuli in Figure 1. Indeed, for all conditions in Figure 2, it is VOT that is the main causative factor, regardless of F0, with perceptual crossovers in the region of the VOT of 20 ms. With hindsight we can say that additional stimuli with VOTs of 15 and 25 ms would have given more precision. At the same time, we do note effects of the fundamental frequency shifts: In each panel there is much spread of data points for 35 ms, and none for 50 ms.

In Figure 3 we focus on the results for the stimuli with a VOT of 20 ms, the one that shows the major effect of F0 shifts. For each of the four F0 onsets we see the percentage of /p/ responses. The coded lines stand for the three durations of F0 shift. A rather general upward trend in /p/ responses is evident as F0 onset rises. A two-way analysis of variance yielded a significant main effect for F0 onset, $F(3, 30) = 36.45$, $p < 0.001$, and a strong interaction between shift-duration and F0 onset for each duration, $F(6, 60) = 6.00$, $p < 0.01$.

Figure 4 focuses on the F0 onset of 130 Hz, the one that had the highest number of /p/ identifications. The /p/ responses for this F0 onset at all four VOT values are shown. Coded lines stand for the three shift durations; the flat F0 plot, marked "no shift," is repeated from Figure 2. It is once again obvious that the major effect is at the VOT of 20 ms, with the deviation from "no shift" increasing with greater shift duration.

The spread of points at the VOT of 5 ms in Figure 4, although much smaller than that at 20 ms, made us look for significant effects in individual cells of the confusion matrix underlying all our plots. That is, wherever we found apparent effects of fundamental frequency at VOT values other than 20, the locus of the main effect, we did a one-tailed t-test for significant deviations from 100%. All such suspicious clusters of responses were at VOT values of 5 ms and 30 ms; for the former, we expected 100% /b/ identifications and for the latter, 100% /p/ identifications. We found three such significant deviations; all of them at the VOT of 5 ms: (1) 120 Hz onset and 50 ms duration, $t(10) = 2.70$, $p < 0.01$, (2) 130 Hz onset and 100 ms duration, $t(10) = -2.51$, $p < 0.025$, (3) 130 Hz onset and 150 ms duration, $t(10) = 2.799$, $p < 0.01$. No such significant deviations were found at the VOT values of 35 ms and 50 ms.

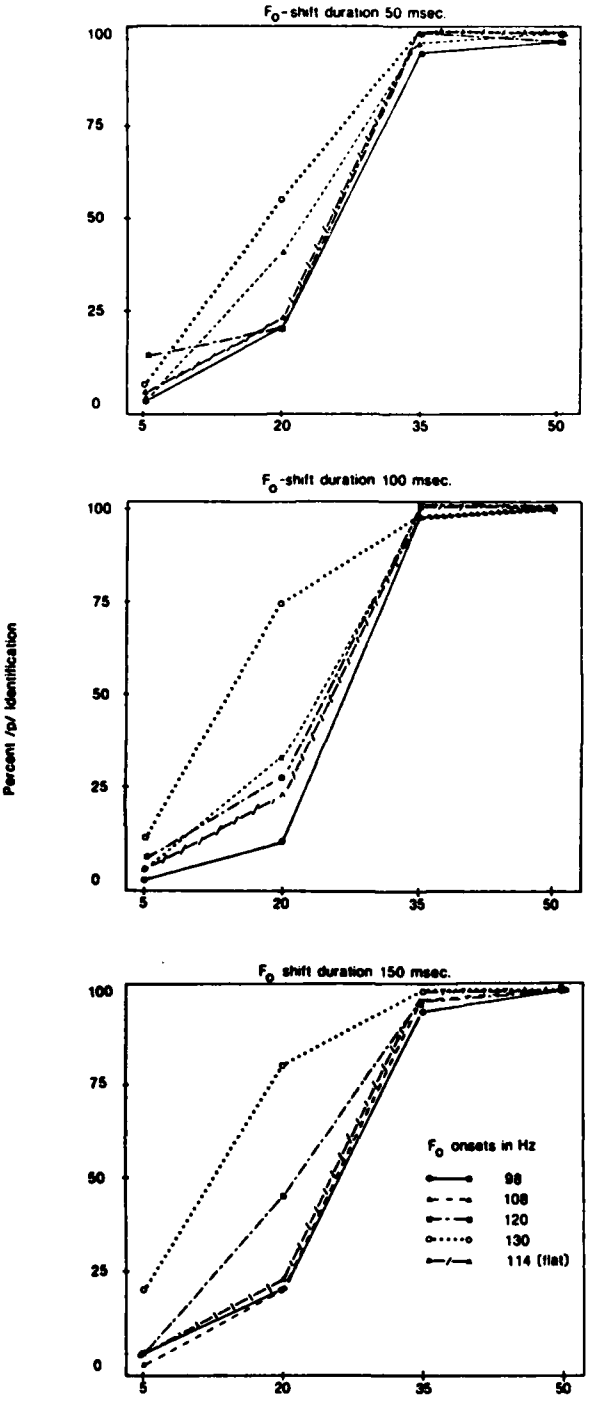


Figure 2. Effects of F0 shifts on identification of VOT variants as English labial stops.

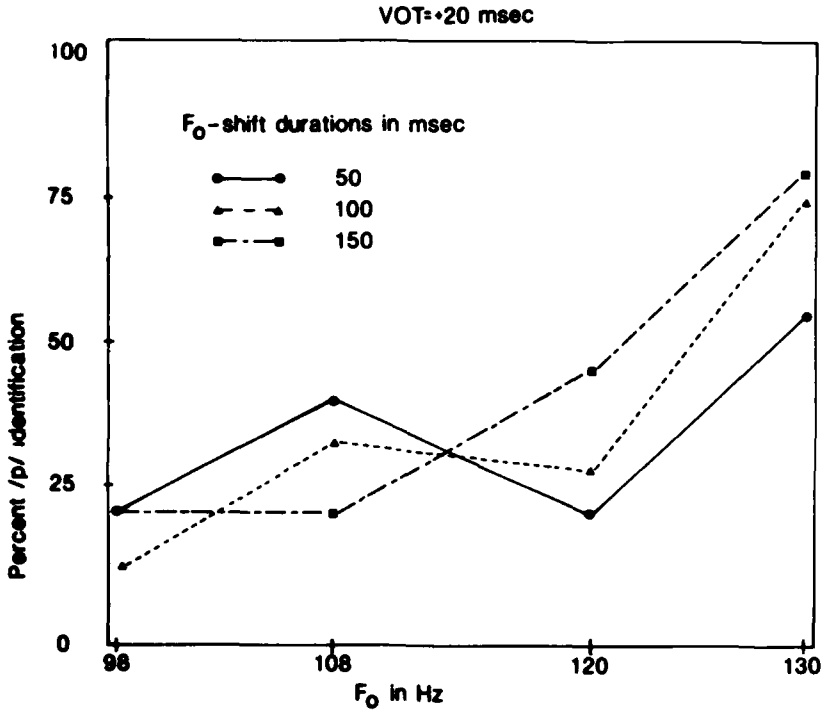


Figure 3. Effects of F0 shifts on VOT of 20 ms.

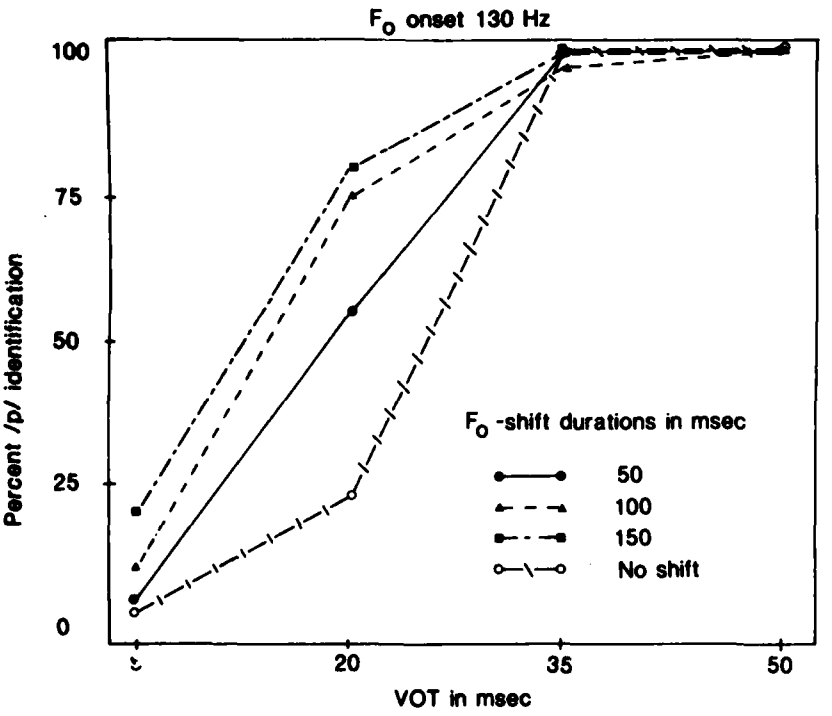


Figure 4. Effects of VOT and shift durations on onset of 130 Hz.

Conclusion

We conclude that there is a modest effect of fundamental frequency shifts on judgments of consonant voicing even within more natural ranges of F0 perturbation[1] than those in Haggard et al. (1970). This is much like the results obtained in the investigation of Thai in an attempt at determining the plausibility of arguments on the rise of distinctive tones (Abramson, 1975; Abramson & Erickson, 1978).

Although they too used a more natural F0 range, Haggard et al. (1981) used an experimental design and stimuli that were somewhat different from ours; their aims were also rather different. To the extent that their data and ours are comparable, they support each other.

If, for the sake of considering the question of relative power of acoustic cues in the perception of a phonetic distinction, we separate fundamental-frequency shifts from the other cues linked to the dimension of voice timing, voice onset time is clearly the dominant cue. Only VOT values that are ambiguous with a flat F0 are likely to be pushed into one labeling category or the other by F0 shifts in a forced-choice test. Finally, there are values of VOT that are firmly categorical; they cannot be affected by F0. There are, however, no values of fundamental frequency that cannot be affected by voice onset time.

References

Abramson, A. S. (1975). Pitch in the perception of voicing states in Thai: Diachronic implications. Haskins Laboratories Status Report on Speech Research, SR-41, 165-174.

Abramson, A. S., & Erickson, D. M. (1978). Diachronic tone splits and voicing shifts in Thai: Some perceptual data. Haskins Laboratories Status Report on Speech Research, SR-53(2), 85-96.

Abramson, A. S., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. Proceedings of the 5th International Congress of Acoustics, Liège: Imp. G. Thone, Paper A51.

Fujimura, O. (1971). Remarks on stop consonants: Synthesis experiments and acoustic cues. In L. L. Hammerich, R. Jakobson, & E. Zwirner (Eds.), Form and substance: Phonetic and linguistic papers presented to Eli Fischer-Jorgensen. Copenhagen: Akademisk Forlag.

Haggard, M. P., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. Journal of the Acoustical Society of America, 47, 613-617.

Haggard, M., Summerfield, Q., & Roberts, M. (1981). Psychoacoustical and cultural determinants of phoneme boundaries: Evidence from trading Fo cues in the voiced-voiceless distinction. Journal of Phonetics, 9, 49-62.

Hombert, J. M. (1975). Towards a theory of tonogenesis: An empirical, physiologically and perceptually-based account of the development of tonal contrasts in language. Unpublished doctoral dissertation, University of California, Berkeley.

Hombert, J. M., Ohala, J., & Ewan, W. (1979). Phonetic explanation for the development of tones. Language, 55, 37-58.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America, 25, 105-113.

Lea, W. (1973). Segmental and suprasegmental influences on fundamental frequency contours. In L. Hyman (Ed.), Consonant types and tone. Southern California Papers in Linguistics (Los Angeles), 1.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. Journal of the Acoustic Society of America, 33, 419-423.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20, 384-422.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia.

Lisker, L., & Abramson, A. S. (1971). Distinctive features and laryngeal control. Language, 47, 767-785.

Maspero, H. (1911). Contribution a l'étude du système phonétique des langues thai. Bulletin de l'Ecole Française d'Extrême-Orient, 19, 152-169.

O'Shaugnessy, D. (1979). Linguistic features in fundamental frequency patterns. Journal of Phonetics, 7, 119-145.

Footnote

[1]The normal ranges of F0 variation linked to consonant voicing, not only in citation forms but especially in running speech (Lea, 1973; O'Shaugnessy, 1979), have still not been well described. We hope to report soon on our current study of this matter with different sentence intonations as a variable.

LARYNGEAL MANAGEMENT AT UTTERANCE-INTERNAL WORD BOUNDARY IN AMERICAN ENGLISH*

Leigh Liskert and Thomas Baer

Abstract. Much attention has been given to the acoustic and physiological means by which the /bdg/-/ptk/ distinction in English is signaled. The most important articulatory difference has been found to involve the nature and timing of laryngeal action associated with the stop articulation. For the labial stops /b/ and /p/, at least three, and possibly four, phonetic classes must be recognized, but we cannot assume that these make up the complete inventory of the ways in which American English speakers coordinate lip and larynx maneuvers in producing these phonemes. Acoustic and physiological data obtained from one American English speaker who produced utterances containing /b/ and /p/ in a variety of contexts showed at least five patterns of lip-larynx coordination, that is, a degree of phonetic versatility usually encountered in studies comparing different speakers across different languages.

Introduction

For many years a good deal of attention has been given to the acoustic and physiological aspects of phonetic distinctions represented by such English word pairs as PILL-BILL, RAPID-RABID, and RIP-RIB. Although the phonetic differences are not precisely the same from pair to pair, we can suppose that they largely reflect differences in the nature and timing of laryngeal adjustments made in association with the closing and opening of the lips. A common effect of these differences is that the first word of each pair is manifested as an acoustic event having a shorter interval of voicing than the second. Since standard phonological analysis and orthography ascribe this voicing difference to one between a phoneme /p/ and a phoneme /b/, it is these phonemes that are characterized as voiceless and voiced, respectively. But while it is enough to posit just two such phonemes in order to provide distinct phonemic spellings of all phonetically different items in the English lexicon that have labial stops, at least three, and possibly four, types of labial stop are generally identified: the phoneme /b/ includes a type with voiced closure and one with voiceless closure, and /p/ has both an aspirated and unaspirated variety of voiceless stops (Gimson, 1962; Trager & Smith, 1951). Moreover, these three or four types may not make up a complete inventory of the ways in which English speakers coordinate laryngeal and supraglottal manuevers when producing utterances that include labial stops; they are at best adequate only for virtually all one-word utterances of the language.

---

*Also Language and Speech, in press. This paper was presented at the 10th International Congress of Phonetic Sciences, 1-6 August 1983, Utrecht.

†Also University of Pennsylvania.

Acknowledgment. Work was supported by NIH grants NS-13617 and NS-13870 to Haskins Laboratories.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

As commonly formulated, the rules that relate phonemic spellings and pronunciation are applicable to single phonological elements, and they have as their domain these entities in certain specified contexts within single words. Thus the phonologist's account of the English labial stops is a set of instructions for pronouncing the letters /p/ and /b/ in hes spellings of English words. But these rules do not provide clear guidance to the pronunciation of /p/ and /b/ in every context in which they are used. In particular, they are silent about lip-larynx management in the case of utterances for which the output of lip and larynx actions is represented by two letters rather than one. Consequently the nature of the difference between the events represented as /b/ + /h/ in ABHOR and RUB HERE and /p/ in APPEAR is not inferable from most accounts of English phonology. Nor can we determine from this literature whether lip-larynx behavior for the forms APPEAR, UPHOLD and STOP HERE are essentially identical or significantly different. If we do not uncritically accept the phonologist's narrow view of phonetic specification as rules for the performance of the letters of hes representation, i.e., if we decline to believe that a phonological spelling cum derivation rules is necessarily the same as a phonetic description, then we may find that the English speakers display a range of systematic variation in lip-larynx coordination considerably greater than is implied by commonly accepted descriptions of the English stop consonants. It may turn out, upon an examination of the kind we describe below, that there is a physical basis, in addition to the well-recognized phonological one, for considering lip and larynx activity in ABHOR, RUB HERE, UPHOLD, and STOP HERE to be gestures for two phonemes in sequence, while in APPEAR those gestures are associated with a single element.

Procedure

In order to gather data giving a more complete picture of lip-larynx relations we made up a list of suitable sentences, as follows:

1. Let's tape each piece separately.
2. Let's play pinochle.
3. Let's just tape hit pieces.
4. Let Abe hit it hard.
5. Did Deb hear what he said?
6. A flip-pistol figured in the heist.
7. Who is Jeb Hill?
8. I don't play billiards.
9. I couldn't help hearing that.
10. I can't tell Pete anything.
11. I think there's a drip here.
12. This is called a drip-pit.
13. Don't trip Bill up.
14. Don't keep pills in your desk.
15. Don't keep bills in your desk.
16. Don't keep hymn books in your desk.
17. Why keep earrings like these?
18. Why keep hearing the same old songs?
19. Why keep peering at your watch?
20. Why keep beer cans in the sink?
21. Is this place light-tight?

22. Is this the right height?
23. Is this side higher?
24. There's a mint here for somebody.
25. Let's have some mint tea with dinner.
26. Let Herb pay the bill.
27. A glib political essay is easy.
28. We'll keep busy till April.
29. They pay plenty for lip service.
30. There's some tape print-through.
31. Let's stay put for a bit.
32. Shooting clay pigeons is great fun.
33. When did the Trib hit the street?

One of us, a speaker of Greater New York City English, read it aloud ten times as the following information was recorded: acoustic waveforms, glotttal aperture as per transillumination (TI), intraoral air pressure, anterior contact, and electromyographic signals from the interarytenoid (INT) and posterior cricoarytenoid (PCA) muscles. We attempted, but failed, to obtain satisfactory signals from the lateral cricoarytenoid. The recorded signals were computer averaged after the ten tokens of each sentence were aligned at the releases of the stops being examined. No other normalizations were imposed.

Results

Figure 1 shows average curves for 500 ms segments excerpted from recordings of the sentences I DON'T PLAY BILLIARDS and LET'S PLAY PINOCHLE. The vertical lines at the midpoints in each panel mark the onsets of the release bursts of the /b/ and /p/ of the words BILLIARDS and PINOCHLE. The curves are, for the most part, just what we should expect: the solid ones for /b/ indicate no change in INT or PCA activity accompanying lip contact, nor is there any sign of glottal opening. The dotted /p/ curves show INT relaxation, PCA contraction, and an opening and closing of the glottis. There are the expected differences in air pressure profiles for /b/ and /p/, as well as differences in the durations of voicelessness or aspiration indicated by the audio envelope curves. More noteworthy is the close similarity of the articulatory contact patterns, which indicates that there is no difference in closure durations.

Figure 2 shows averaged data for three sentences (#17,18,19), the relevant phrases being KEEP EARRINGS, KEEP PEERING, and KEEP HEARING. The word-final /p/ before the vowel in KEEP EARRINGS was produced with no apparent glottal opening during the interval of labial contact and elevated air pressure, although there was INT slackening and some PCA contraction. (For some tokens of this sentence, the word-initial vowel was glottalized at onset.) The picture for KEEP PEERING is very like the one for the /p/ of PINOCHLE shown in Figure 1. The similarity amounts to identity in the transillumination profiles, although the PCA and pressure signals are high for a longer time in KEEP PEERING. Note that although INT and PCA adjustments begin in time with the onset of the long closure of KEEP PEERING, the peak of glottal opening is as closely synchronized with the release as in the case of the simple aspirated /p/ of PINOCHLE.

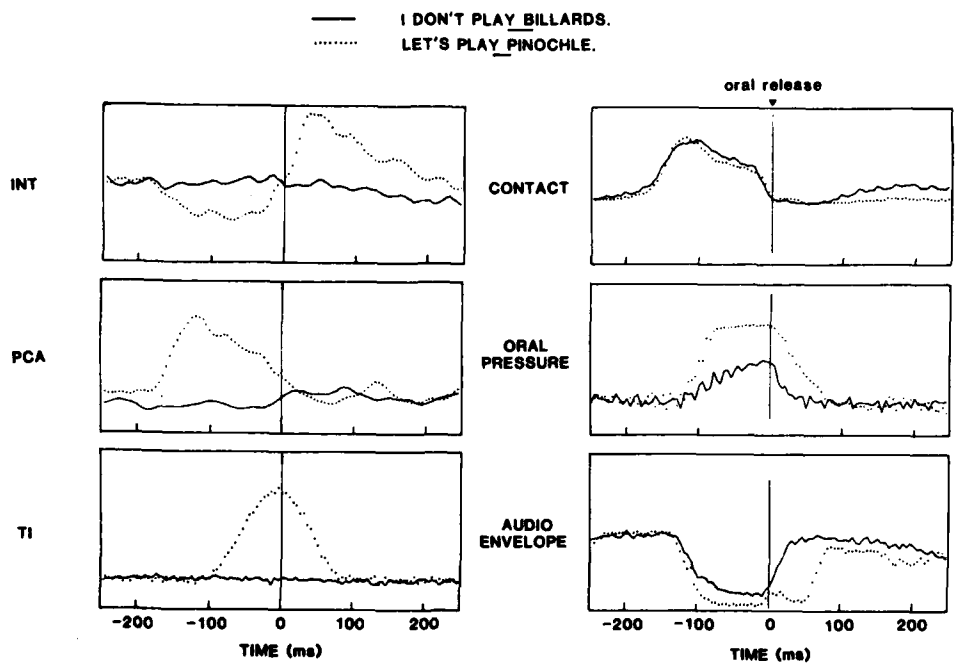


Figure 1. Activity associated with the production of bilabial stops in the sentences LET'S PLAY PINOCHLE and I DON'T PLAY BILLIARDS. Electromyographic, transillumination, articulatory contact, introral air pressure, and audio data are shown. Each curve is an ensemble average calculated from ten tokens of each sentence. Ordinate scales have been omitted to simplify the figure. Zero time represents oral release for the underlined stops.

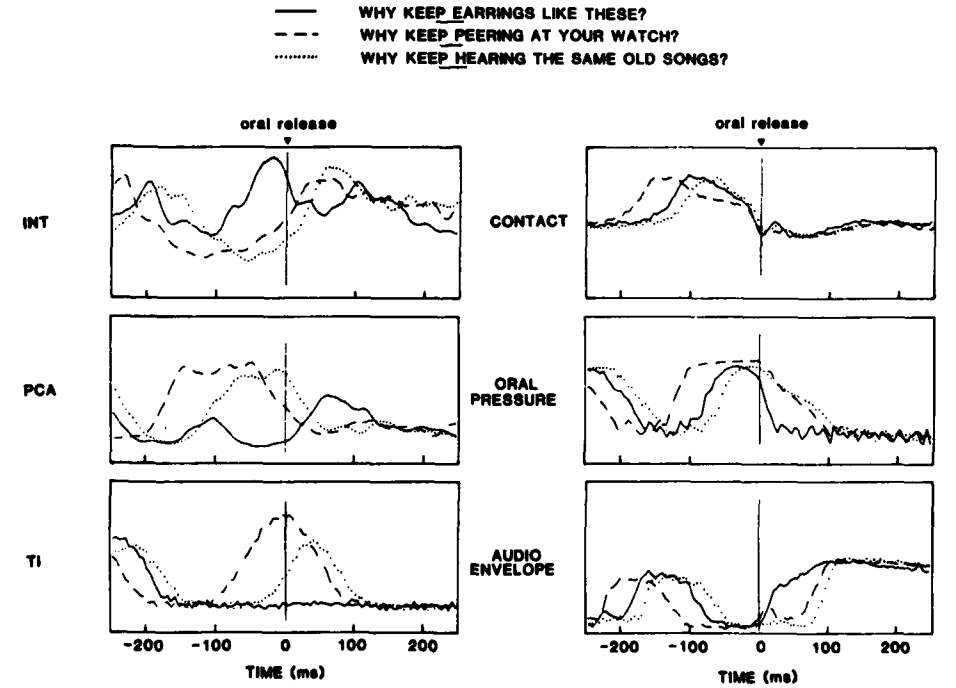


Figure 2. Averaged data for word-final /p/ in three different sentence environments.

In KEEP HEARING (the dotted curves of Figure 2) the beginning and peak of glottal opening are about 40 ms later than in KEEP PEERING (the dashed curves), and this is presumably to be connected with the difference in the audio signal profiles, which suggests that voicing resumes later in KEEP HEARING. This greater lag in the resumption of voicing is clearly not to be explained by a greater glottal opening at the time of release, or by a greater magnitude of peak opening.

The finding that the word-initial aspirated /p/ is released when glottal aperture is maximum, as in PINOCHLE (Figure 1) and KEEP PEERING (Figure 2) turns out, upon further examination of our data, not to hold generally for this stop type. In Figure 3 the solid curves represent transillumination data for the sentences LET'S PLAY PINOCHLE, THIS IS CALLED A DRIP-PIT, and I DON'T PLAY BILLIARDS. The aspirated /p/s following LET'S, DRIP, and DON'T were all released after the point of maximum glottal aperture was past. The glottal aperture of LET'S PLAY is no doubt as much associated with the /s/ as with the /p/, which may explain why it is early relative to the release. Perhaps in all three sentences there is something about their prosodies that is a factor in advancing the time of glottal opening and closing, but it is nevertheless puzzling that the /p/ of LET'S PLAY is well aspirated though the glottis at release is already two-thirds the way to closure. This result is especially puzzling in the light of other published data on /s-k/ sequences (Yoshioka, Löfqvist, & Hirose, 1981) and /s-t/ sequences (Pètursson, 1978) with intervening word boundaries that show a second peak of glottal opening centered at the release of the stops.

When we compare sentences said to involve /b/+/h/ and /p/+/h/ sequences, as per Figure 4, we find little difference in contact patterns, in transillumination profiles, or in the time at which the audio signals return to full amplitude after the stop releases. The only difference in glottal aperture patterns is the voicing ripple for the sequence with /b/ in contrast to the smooth curve for /p/; the temporal courses and magnitudes of opening are precisely the same. The INT and PCA patterns are also very much alike for the two sequences. We note, of course, the expected differences in oral pressure.

Summary

Our data appear to bear out the truth of the supposition motivating the experiment just reported--namely, that a description of lip-larynx coordination patterns limited to the /p/-/b/ contrast in such word pairs as PILL-BILL, RAPID-RABID, and RIP-RIB fails to account for all the patterns to be found in English. In all, at least so far, as many as five may be enumerated: 1) Intervocalic /b/ is produced with no change in the settings of the INT and PCA muscles or in the glottal aperture appropriate to the neighboring vowels. 2) The unaspirated /p/ in intervocalic position is accomplished with no discernible opening of the glottis, although there is some PCA contraction and INT relaxation. 3) Sequences of word-final voiceless obstruent and aspirated /p/ are produced with the PCA and INT adjustments that serve to open the glottis, the peak of this opening being variable and ranging from as early as 100 ms to just slightly before release. 4) An aspirated /p/ following a vowel, but not in word-final position, is produced with a glottal opening that peaks in close synchrony with the stop release, 5) Signal intervals interpreted as a labial stop followed by the phoneme /h/ show glottal openings that peak well after the release (VOT=+50 ms), with the salient difference between /b/+/h/ and

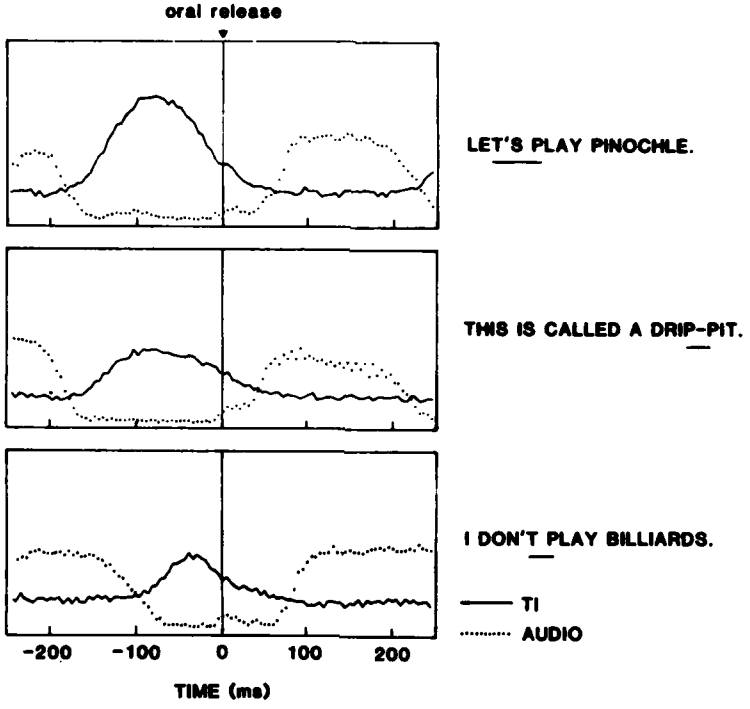


Figure 3. Averaged transillumination and audio data for word-initial /p/ following voiceless consonants in three sentences.

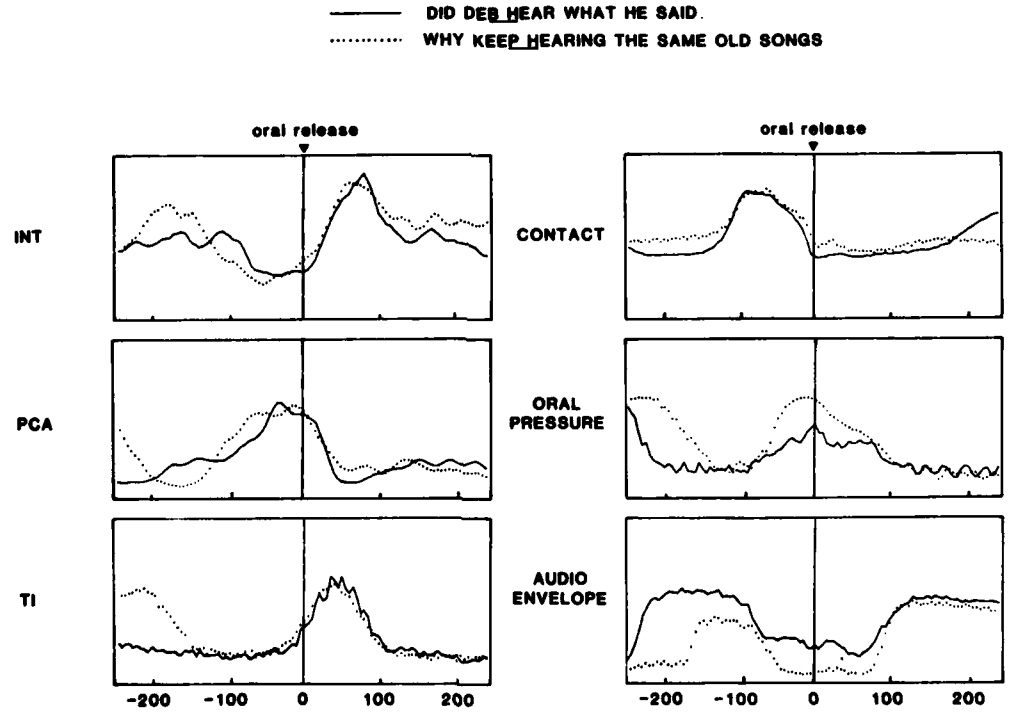


Figure 4. Averaged data for word-final /b/and /p/ before /h/.

/p/+/h/ a matter of voicing over the combined intervals of oral closure and glottal opening, despite the absence of any observable difference in INT and PCA behavior.

Concluding Comment

The observed differences in glottal aperture profiles in relation to supraglottal events cannot be entirely understood on the basis of our EMG data, a fact that is not surprising in view of the limitations of this study. It is generally agreed that while the PCA may be the only abductory muscle, the lateral cricoarytenoid (LCA) and thyroarytenoid (TA) muscles as well as the INT muscle play a role in vocal fold adduction, and hence in determining the extent to which PCA contraction is effective in opening the glottis (Sawashima & Hirose, 1983). We may therefore reasonably suppose that, had we managed to tap one or more additional muscles of the larynx, we would be better able to explain the apparent anomalies in the data on the type 2, 3, and 5 patterns. Thus we might account for the finding that the unaspirated /p/ is produced without glottal opening although INT and PCA signals favor it. This finding is in agreement with Dixit's (1975) description of the Hindi voiceless unaspirated stops, and at variance with the results reported by Benguerel and Bhatia (1980). English speakers show considerable variability in the frequency and degree to which such stops are "glottalized" (as judged auditorily) and accompanied by separation of the arytenoid cartilages (Sawashima, 1970), and it is possible that Hindi speakers are as free with this feature as English speakers. EMG data reported both by Hirose, Lisker, and Abramson (1977) and Dixit (1975) indicate that data on the LCA and TA muscles would resolve the apparently contradictory findings. Such information, in addition, would possibly tell us how the voicing difference between /p/+/h/ and /b/+/h/ (Figure 4) is managed without any apparent difference in PCA and INT activity.

As was said earlier, the greater duration of aspiration for /p/+/h/ than for aspirated /p/ cannot be explained, as per Kim (1970), by a greater magnitude of glottal aperture at release, but rather by the longer delay of the laryngeal gesture relative to the labial release. At release the aspirated /p/ has the greater aperture, but the glottis begins to close at that time; the glottis is less open at the release of /p/ before /h/, but it is still increasing in aperture. This may explain not only the difference in the duration of aspiration, but also our auditory impression, one consistent with a difference in their waveforms, that the release burst and the aspiration for /p/+/h/ are both of weaker intensity.

Finally, it may be of some phonological interest that the degree of overlap in lip-larynx activity is greater for the voiceless stop plus aspiration that is interpreted as a single element than for those represented phonologically as /p/+/h/ and /b/+/h/. It is tempting to infer from this that the phonologist's decision as to whether one or two elements are involved is phonetically based, but a comparison of our data with those reported for Hindi by Dixit and by Benguerel and Bhatia forces us to recognize that the decision is primarily dictated by morphosyntactic considerations. It is true that English /p/+/h/ and Hindi /ph/, which may well be produced with equal delays in voice onset, differ in that peak glottal opening is later for the English two-phoneme sequence; English /b/+/h/ and Hindi /bh/, however, show no similar difference to justify a claim that their different phonological status derives from a phonetic difference. The basis for denying that English possesses voiced aspirated stops and voiceless stops of two degrees of aspira-

tion is not phonetic at all. At the same time it can be said that phonetic data of the kind presented above provide ancillary support for the phonological distinction made between aspiration as one of the features of /p/ and as an independent phonological element /h/ that freely occurs after a large number of other elements, including /p/ and /b/.

References

Benguerel, A-P., & Bhatia, T. K. (1980). Hindi stop consonants: an acoustic and fiberscopic study. Phonetica, 37, 134-148.

Dixit, R. P. (1975). Neuromuscular aspects of laryngeal control: with special reference to Hindi. Unpublished doctoral dissertation, University of Texas, Austin.

Gimson, A. C. (1962). An introduction to the pronunciation of English. London: Edward Arnold.

Hirose, H., Lisker, L., & Abramson, A. (1977). Physiological aspects of certain laryngeal features in stop production. Haskins Laboratories Status Report on Speech Research, SR-31/32, 183-191.

Kim, C-W. (1970). A theory of aspriation. Phonetica, 21, 107-116.

Pêtursson, M. (1978). Jointure au niveau glottal. Phonetica, 35, 65-85.

Sawashima, M. (1970). Glottal adjustments for English obstruents. Haskins Laboratories Status Report on Speech Research, SR-21/22, 187-200.

Sawashima, M., & Hirose, H. Laryngeal gestures in speech production. In P. F. MacNeilage (Ed.), The production of speech (pp. 11-38). New York: Springer-Verlag.

Trager, G. L., & Smith, H. L. Jr. (1951). An outline of English structure. (Studies in Linguistics: Occasional Papers, 3). Norman, OK: Battenburg Press.

Yoshioka, H., Löfqvist, A., & Hirose, H. (1981). Laryngeal adjustments in the production of consonant clusters and geminates in American English. Journal of the Acoustical Society of America, 70, 1615-1623.

CLOSURE DURATION AND RELEASE BURST AMPLITUDE CUES TO STOP CONSONANT MANNER AND PLACE OF ARTICULATION*

Bruno H. Repp

Abstract. The perception of stop consonants was studied in a constant neutral [s-l] context. Truncated natural [p], [t], and [k] release bursts at two intensities were preceded by variable silent closure intervals. The bursts, though spectrally distinct, conveyed little specific place information but contributed to the perception of stop manner by reducing the amount of silence required to perceive a stop (relative to a burstless stimulus). Burst amplitude was a cue for both stop manner and place; higher amplitudes favored t, lower amplitudes favored p responses. The silent closure interval, a major stop manner cue, emerged as the primary place cue in this situation: Short intervals led to t, long ones to p responses. All these perceptual effects probably reflect listeners' tacit knowledge of systematic acoustic differences in natural speech.

Silent closure duration is an important cue to the perception of stop consonant manner--that is, of phonetic distinctions that rest on the perceived presence versus absence of a stop consonant (e.g., Bailey & Summerfield, 1980; Dorman, Raphael, & Liberman, 1979; Repp, 1984). The question of principal interest in the present study was whether different amounts of closure silence are needed to perceive stop consonants having different places of articulation. Specifically, it was hypothesized that, because labial stops generally have longer closure durations than alveolar and velar stops in natural speech (e.g., Bailey & Summerfield, 1980; Menon, Jensen, & Dew, 1969; Stathopoulos & Weismer, 1983; Suen & Beddoes, 1974), longer intervals might be needed for their perception, too.

This hypothesis makes two semi-independent predictions: (1) Given unambiguous cues to stop consonant place of articulation, more silence will be needed to perceive p than t or k; that is, perception of stop manner, as cued by closure duration, may depend on perceived place of articulation. (2) Given ambiguous place cues and sufficient silence to perceive a stop consonant, short closure silences will lead to t or k responses while long silences will lead to p responses; that is, closure duration is a direct cue to place of articulation. The first of these predictions is difficult to test because the different acoustic configurations needed to specify place of articulation unambiguously may have psychoacoustic effects on perception of the closure

---

*Also Language and Speech, in press.
Acknowledgment. This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. A short version of this paper was presented at the 105th meeting of the Acoustical Society of America in Cincinnati, Ohio, May 1983.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

silence, which are difficult to dissociate from phonetic effects due to perceived place of articulation. The second prediction, however, can be tested easily by varying silence duration in a constant acoustic environment.

In a previous study addressing these issues, Bailey and Summerfield (1980) used synthetic speech stimuli consisting of an initial [s] noise followed by a variable silent interval and by a vocalic portion with or without initial formant transitions. Two findings are relevant here. When either the second formant of a steady-state vowel or the vocalic formant transitions were varied so as to cue the perception of p, t, or k unambiguously, the amount of silence required to perceive the stop consonant did not vary significantly with place of articulation, except that it was reduced for k cued by formant transitions. Bailey and Summerfield attributed this latter effect to auditory energy summation caused by the proximity of the second and third formants at vowel onset; that is, they assumed a psychoacoustic rather than phonetic basis for the effect. The other finding was that, when the place-of-articulation cues in the vocalic portion were ambiguous, so that (given sufficient silence) the same acoustic pattern elicited more than one type of stop response, p responses were clearly preferred at longer closure durations, while t or k responses predominated at short closures. The first, negative finding suggests that stop manner perception is largely independent of perceived place of articulation. The second finding, however, suggests that the listeners' internal perceptual criteria for place of articulation do include closure duration as an important acoustic dimension.

The principal aim of the present study was to replicate Bailey and Summerfield's findings, using natural-speech stimuli that, instead of variable formant frequencies or transitions, included release bursts appropriate for each place of articulation. A second aim was to examine the specific contribution of the release burst itself to stop manner perception. As a rule, alveolar and velar stops following [s], in contrast to labial stops, do not need any closure silence to be perceived as long as an intact natural release burst is present (Repp, 1984). This difference in silence requirements might be due to the higher amplitude and longer duration of alveolar and velar bursts (Zue, 1976), and it might disappear when the overall amplitudes of these bursts are reduced to resemble those of labial bursts. In addition to examining this question, the present experiment also investigated to what extent burst amplitude affects perception of stop manner and place, following Ohde and Stevens (1983) and Repp (1984).

Two methodological decisions require justification. First, to exclude cues to stop place of articulation in the signal portions surrounding the critical cues of closure duration and release burst, these cues were embedded in a constant [s-l] context. Preliminary observations suggested that [l] resonances contain only weak (if any) formant transition cues to preceding stop consonants, so this segment seemed ideally suited for the purpose. However, this resulted in some consonant clusters ([stl] and [skl]) that are unfamiliar to English speakers and listeners. It was assumed, however, that these clusters would not be difficult to produce or perceive, and the results tend to justify this assumption. Second, in order to make closure duration a salient cue to stop manner at all three places of articulation, it was necessary to reduce the natural release bursts, since full alveolar and velar release bursts are generally sufficient cues for perception of a stop consonant. This was done by waveform truncation and resulted in residual bursts that were spectrally distinct but, as it turned out, conveyed surprisingly little place

information. The present study thus primarily addresses the question of the role of closure duration as a cue when other place-of-articulation cues are highly ambiguous.

Method

Stimuli

A number of repetitions of the utterances slat, splat, stlat, and sclat were recorded by a male speaker of American English, low-pass filtered at 4.8 kHz, and digitized at 10 kHz. One good token of each utterance was selected and manipulated further by computer waveform editing procedures. The release bursts (i.e., the aperiodic signal portion preceding the first glottal pulse) of splat, stlat, and sclat (originally 17, 43, and 43 ms in duration, respectively) were excerpted and trimmed to 10 ms duration. This was done by eliminating the final low-amplitude portions of the labial and alveolar bursts. The velar burst, on the other hand, had several amplitude peaks, the last and most pronounced of which happened to occupy the last 10 ms; therefore, this final portion was taken as the truncated burst. Two versions of each truncated burst were created by changing their amplitudes by 10 dB: The labial burst was amplified by that amount while the alveolar and velar bursts were attenuated. This was done because the labial burst had less high-frequency energy than the other two bursts (see below). Each of these six bursts was spliced onto the lat portion (365 ms long) derived from slat; thus, the voiced portion immediately following each burst was constant and contained no distinctive cues to place of stop articulation. A seventh, burstless stimulus was included as a baseline. All seven stimuli were preceded by a constant [s]-noise (226 ms long) derived from slat, and by a variable closure interval. Closure intervals were varied from 0 to 100 ms in 20-ms steps, for a total of 35 stimuli that were recorded in 5 different random orders.

Subjects and Procedure

Ten subjects (nine paid student volunteers and the author) listened to the stimulus tape over TDH-39 earphones at a comfortable intensity (approximately 76 dB SPL for vowel peaks) and identified the stimuli in writing as beginning with sl, spl, stl, or scl. Instructions alerted subjects to the unfamiliar consonant clusters.

Results and Discussion

Figure 1 compares the labeling function (percent stop responses, regardless of place of articulation, as a function of closure duration) for burstless stimuli with the average labeling function for the six types of stimuli with bursts. As indicated in the figure by the horizontal bar at the 50-percent point, the average phonetic boundaries for the six burst conditions varied over a 10-ms range, from 34.5 to 44.5 ms of closure silence. The boundary for the burstless stimuli was clearly longer--at a nominal 50.5 ms of silence (i.e., measured to the onset of the nonexisting burst), or at an actual 60.5 ms of silence (as indicated by the arrows in the figure). This difference was exhibited by all subjects and was significant in a one-way analysis of variance on the total percentage of stop responses, after applying a correction for the conversion to nominal closure duration and after omitting the data for the author who showed the largest difference, $F(1,8) = 16.6$, $p < .01$. Thus, the truncated release bursts made a significant contribution to stop manner perception (cf. Repp, 1984); that is, the boundary was shortened by more than

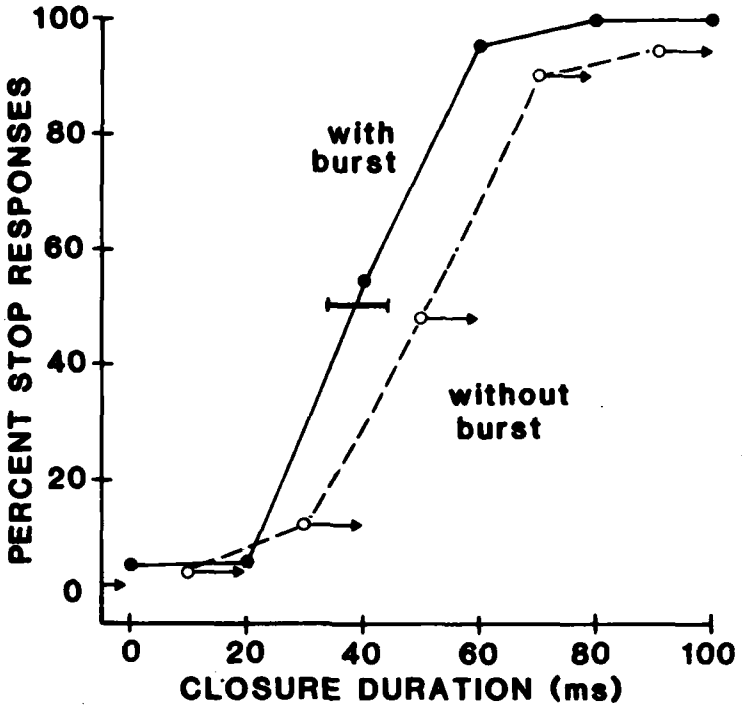


Figure 1. Effect of presence versus absence of release burst. The solid function is the average of all six burst conditions; the horizontal bar indicates the range of 50-percent cross-overs. Closure durations in the no-burst condition are nominal; actual durations are indicated by arrows.

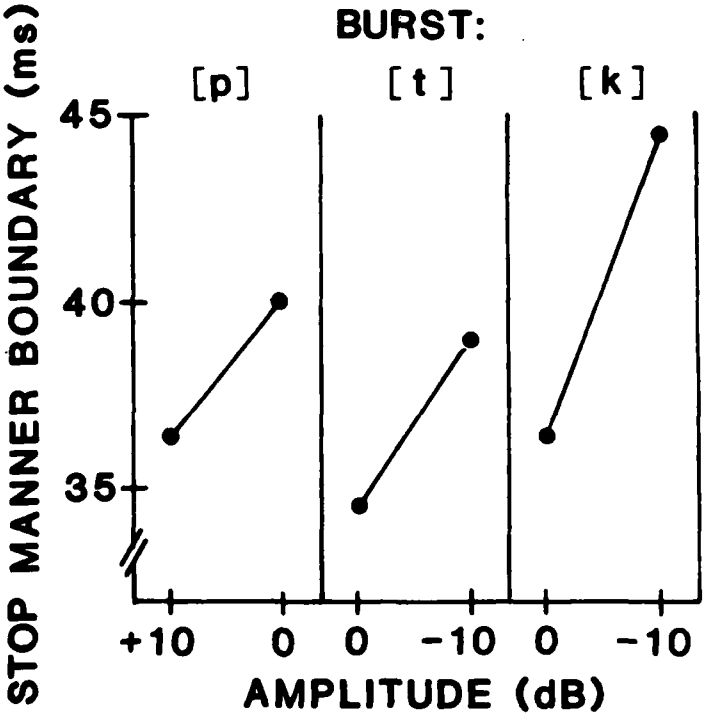


Figure 2. Comparison of category boundaries in six burst conditions: Effects of burst category and amplitude.

the 10 ms expected if the presence of a burst merely had prolonged the effective closure duration.

Figure 2 shows the effects of burst category (intended place of articulation) and amplitude on the stop manner boundary, still combining all kinds of stop responses. Burst amplitude clearly had an effect: Amplification of the labial burst increased stop responses (i.e., shortened the boundary) while attenuation of the alveolar and velar bursts decreased stop responses. Thus, burst amplitude could be traded against closure silence in stop manner perception (cf. Repp, 1984). The effect of burst amplitude was significant in an analysis of variance, $F(1,9) = 8.4$, $p < .05$. The main effect of burst category was nonsignificant, and so was the interaction.

A comparison across the three burst categories is difficult because amplitude differences are confounded with spectral differences. Overall rms amplitudes were determined after redigitizing the stimuli without preemphasis. Unexpectedly, the amplitude of the labial burst turned out to be 3 dB higher than that of the alveolar and velar bursts, which were equal and 6 dB below the amplitude of the [l] onset (the first 10 ms). This was apparently due to a strong low-frequency component in the labial burst waveform. It is likely, however, that the amplitude of higher-frequency components is more important for stop manner perception, as has also been hypothesized by Ohde and Stevens (1983) with regard to place of articulation perception. Figure 3 compares the spectra of the three truncated bursts at their original amplitudes. As expected, the labial burst had less energy than the alveolar and velar bursts in the high-frequency regions above 2 kHz; the average difference is about 10 dB. Thus, amplification of the labial burst by 10 dB resulted in approximately equal levels of high-frequency energy across the three burst categories, which is consistent with the very similar stop manner boundaries obtained (see Figure 2).

So far, stop responses have been treated as a single category. We turn now to an analysis of stop responses by place of articulation. Figure 4 shows conditional percentages of p, t, and k (i.e., scl) responses in separate panels as a function of closure duration (from 40 ms up) and of burst category, combining the two burst amplitudes. The no-burst condition is also plotted at the actual closure durations. It is evident that closure duration provided the most important cue to stop place of articulation: At short closures, t responses predominated (notwithstanding a possible bias against reporting stl clusters) while, at long closure durations, the response was overwhelmingly p. These trends held almost regardless of the nature of the burst; [p] and [t] bursts, in particular, yielded highly similar results. The results for [k] bursts resembled those for burstless stimuli, perhaps because this late component of the burst did not preserve specific place of articulation information. (Cf. the absence of a pronounced mid-frequency peak in the spectrum--see Figure 3--which characterizes velar onset spectra, according to Stevens and Blumstein, 1978.)

The similar perceptual results for [p] and [t] bursts, whose spectra did exhibit the general spectral properties characteristic of these places of articulation (see Figure 3 and Stevens and Blumstein, 1978; note that the present spectra are not pre-emphasized), may have been due to their short duration. According to Stevens and Blumstein (1978), the most salient place cue is the onset spectrum computed over a window approximately 25 ms long; if so, the 10-ms bursts were presumably integrated with the constant [l] onset

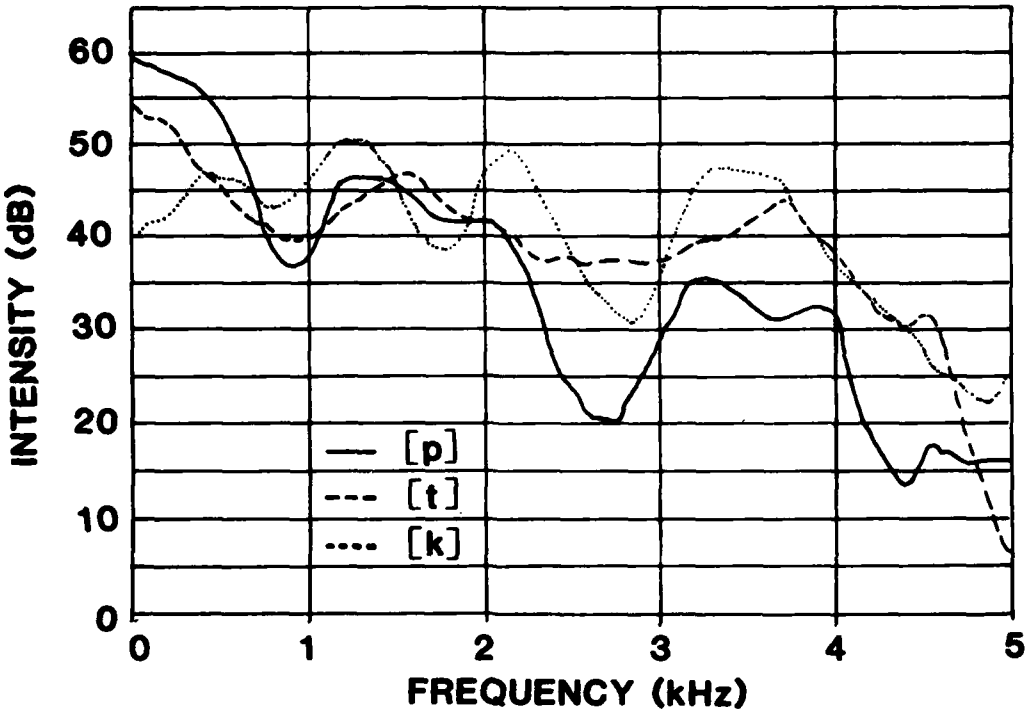


Figure 3. Spectra of the 10-ms [p], [t], and [k] bursts at their original amplitudes, without pre-emphasis. Spectra were obtained by FFT analysis (program FDI of the ILS package), using a 25.6-ms Hamming window whose left edge preceded the burst onset by 10 ms. The spectra were smoothed by averaging across a 400-Hz rectangular window moving in 20-Hz steps.

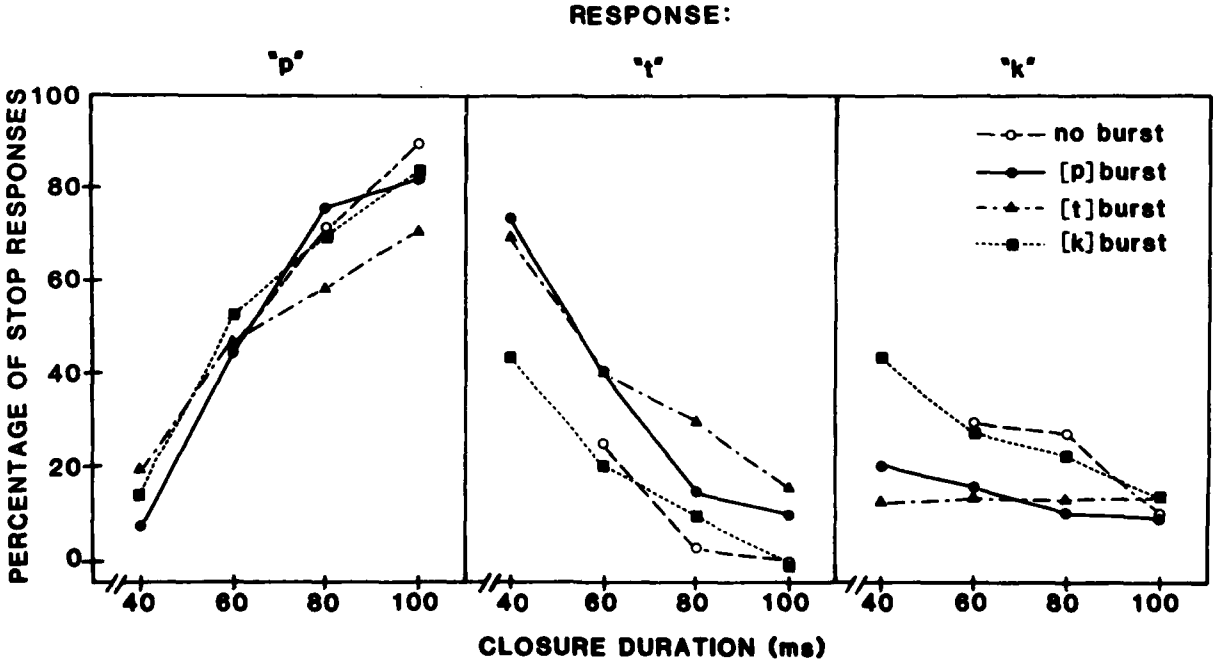


Figure 4. Percentages of p, t, and k responses (separate panels) as a function of closure duration and burst condition.

following them and thus lost much of their distinctiveness. The present results show very clearly, however, that more than the onset spectrum is involved in place perception: When spectral cues are ambiguous, closure duration takes over as the salient place cue, as also observed by Bailey and Summerfield (1980).

The reason for the effectiveness of the closure duration cue presumably lies in the well-known fact that [p] closures tend to be longer in natural speech than [t] and [k] closures (although little is known about [stl] and [skl] clusters).1 An alternative, psychoacoustic explanation might be proposed, however: that the preceding [s]-noise, with its strong high-frequency components, left a trace in sensory memory that was integrated with the onset spectrum following a short closure. Such integration might explain the bias toward t responses at short closures, assuming that the predominating response after removal of the preceding [s]-noise would be p (or, rather, b). Even though research on adaptation in the auditory nerve (e.g., Delgutte & Kiang, 1984) predicts spectral contrast rather than integration, a brief additional test was conducted to address this question. Ten randomized repetitions of the seven stimuli (six with bursts and one without) without the initial [s]-noise were presented for identification as lat, blat, dlat, or glat to a new group of nine subjects plus the author. The results were mixed. Two subjects responded randomly. Four subjects identified the burstless stimulus as lat but labeled all others predominantly blat. The remaining four subjects (including the author) distributed their responses more evenly, although accuracy was poor (45 percent correct for stimuli with bursts; 100 percent for lat). These results show, first, that the relative ineffectiveness of the bursts as place cues in the present experiment was not due to the preceding [s]-noise and closure. Second, although some subjects showed a strong bias toward b responses, this bias was not so universal as to lend convincing support to the hypothesis that the striking change from t to p responses with increasing closure duration in the main experiment was due to spectral integration. More likely, the effect of closure duration has a phonetic origin. That is, listeners expect labial stops to have longer closures on the basis of their knowledge of natural speech patterns.

Finally, Figure 5 provides a different breakdown of the data, which reveals effects of burst amplitude on perceived place of articulation. The conditional percentage of responses in each stop category, averaged over closure durations from 40 to 100 ms, is shown for each of the six burst category/amplitude conditions. "Correct" responses (i.e., responses reflecting the place of articulation that the burst was intended for) are indicated by the cross-hatched bars. It is evident that correct responses in each stop category decreased as burst amplitude was modified, due to a higher percentage of p responses for weak bursts and of t and/or k responses for strong bursts. This result replicates earlier findings of Ohde and Stevens (1983) with synthetic speech. Despite the relative weakness of the present bursts as specific place cues, it appears that burst amplitude contributed to place as well as manner perception.

Conclusions

The present findings are consistent with many other results suggesting that listeners possess detailed tacit knowledge of the acoustic correlates of phonetic categories (see Repp, 1982, for a review). The perceptual criteria derived from this knowledge apparently specify that labial stops ought to have

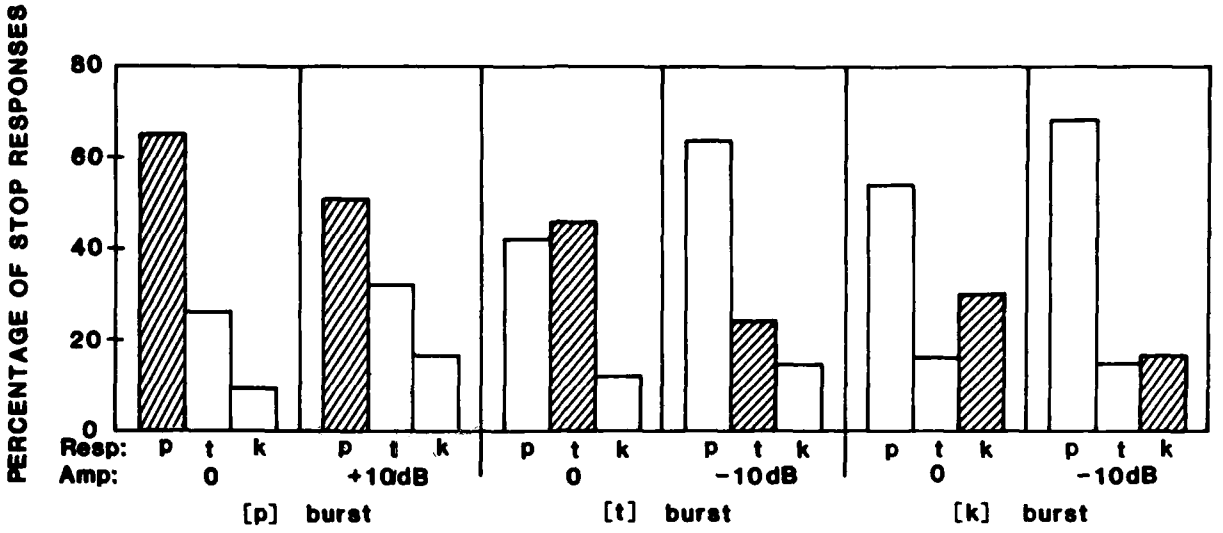


Figure 5. Response distributions in the six burst conditions, averaged over closure durations.

a longer closure interval than alveolar or velar stops. They also specify that labial stops ought to have weaker release bursts; hence the effect of burst amplitude on place of articulation perception. These perceptual criteria presumably derive from experience with natural speech in its acoustic and articulatory manifestations, and they provide the frame of reference within which speech perception takes place.

References

Bailey, P. J., & Summerfield, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. Journal of Experimental Psychology: Human Perception and Performance, 6, 536-563.

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve. IV. Sounds with consonant-like dynamic characteristics. Journal of the Acoustical Society of America, 75, 897-907.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. (1979). Some experiments on the sound of silence in phonetic perception. Journal of the Acoustical Society of America, 65, 1518-1532.

Menon, K. M. N., Jensen, P. J., & Dew, D. (1969). Acoustic properties of VCV utterances. Journal of the Acoustical Society of America, 46, 449-457.

Ohde, R. N., & Stevens, K. N. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. Journal of the Acoustical Society of America, 74, 706-714.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. Psychological Bulletin, 92, 81-110.

Repp, B. H. (1984). The role of release bursts in the perception of [s]-stop clusters. Journal of the Acoustical Society of America, 75, 1219-1230.

Stathopoulos, E. T., & Weismer, G. (1983). Closure duration of stop consonants. Journal of Phonetics, 11, 395-400.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. Journal of the Acoustical Society of America, 64, 1358-1368.

Suen, C. Y., & Beddoes, M. P. (1974). The silent interval of stop consonants. Language and Speech, 17, 126-134.

Zue, V. W. (1976). Acoustic characteristics of stop consonants: A controlled study. Lincoln Laboratory Technical Report No. 523 (Lexington, MA).

Footnote

[1]The speaker of the utterances for this experiment, an experienced linguist, produced five tokens each of splat, stlat, and sclat. Average closure durations were 111, 112, and 68 ms, respectively, revealing unusually long values for [t]. For spat, stat, scat, and for sprat, strat, scrat, produced by the same speaker, however, closure durations ranked [p] > [k] > [t]. Clearly, more data are needed to determine whether [-l] context is an exception to the rule that labial closures are longest in duration.

EFFECTS OF TEMPORAL STIMULUS PROPERTIES ON PERCEPTION OF THE [sl]-[spl] DISTINCTION*

Bruno H. Repp

Abstract. Two studies investigated the influence of the independently varied durations of preceding and following signal portions on the amount of closure silence needed to perceive splash rather than slash. Increases (or decreases) in the durations of the [s] and [l] acoustic segments had opposite effects that cancelled when the silent intervals were short (Exp. 1), but yielded a net effect due to [s] duration when the silent intervals were long (Exp. 2). These findings, which resolve a conflict between earlier results in the literature, are interpreted as reflecting a perceptual compensation for coarticulatory shortening of [s] before stop consonants, in conjunction with (possibly psychoacoustic) contrastive interactions between the perceived durations of adjacent acoustic segments. The results suggest that local temporal signal properties, as distinct from global perceived speaking rate, are an important factor in phonetic perception.

An important perceptual cue for the distinction between the word-initial clusters [sl] and [spl] is the absence versus presence of a silent interval following the [s] noise (e.g., Bastian, Eimas, & Liberman, 1961; Fitch, Halwes, Erickson, & Liberman, 1980). Two fairly recent studies have investigated whether the category boundary on a continuum ranging from slit to split, created by varying the duration of the silent closure interval, is affected by reductions in total stimulus duration: Marcus (1978) found that temporal compression left the slit-split boundary unaffected, whereas Summerfield, Bailey, Seton, and Dorman (1981) found that less silence was needed to perceive split in temporally compressed stimuli. Both studies made use of modified natural-speech tokens of slit; Summerfield et al. also used synthetic stimuli, with similar results. In an attempt to explain the difference in outcomes, Summerfield et al. pointed out that the category boundaries in the Marcus study were at considerably shorter silences (less than 30 ms) than the boundaries in their own study (around 60 ms). They conjectured (as had Marcus) that a perceptual limit, perhaps related to an articulatory limit, may be encountered at short silences, and that this may be the reason why the boundary refused to shift to even shorter values in the Marcus study. They

---

*Also Phonetica, in press.

Acknowledgment. This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. A short version of this paper was presented at the 107th meeting of the Acoustical Society of America in Norfolk, Virginia, May 1984. I am grateful to Peter Bailey, Joanne Miller, Richard Pastore, and especially D. H. Whalen for helpful comments on an earlier draft.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

interpreted their own findings as reflecting a perceptual adjustment to variations in contextual speech rate.

The principal reason for conducting the present experiments was the author's suspicion that temporal changes in signal portions preceding and following the silence may not be equally relevant. Several earlier perception studies in which closure silence duration was the dependent variable, albeit for different phonetic contrasts, have found that the duration of the preceding acoustic segment has a much stronger effect than that of the following segment (Port & Dalby, 1982; Repp, 1979). In addition, there is another reason to expect [s] duration to be important, quite regardless of perceived speaking rate: Fricative noise duration tends to be shorter in [spl] than in [sl] clusters (Morse, Eilers, & Gavin, 1982; Schwartz, 1969, 1970; see also Haggard, 1973), and listeners may have tacit knowledge of this coarticulatory relationship, as they do of so many others (see Repp, 1982). The duration of [l], on the other hand, does not seem to exhibit such coarticulatory variation (Morse et al., 1982; Repp, unpublished data) and therefore may be perceptually irrelevant. To examine this hypothesis, the durations of the fricative noise and of the lateral resonance were varied independently in the present experiments.

Experiment 1

Experiment 1 used a slash-splash continuum (from Repp, 1984: Exp. 7) for which the average category boundary happened to be around 25 ms of silence, similar to the short boundary obtained by Marcus (1978). This provided an opportunity to test further the hypothesis of a lower limit for the perception of silence duration in this context. While the reason for the short boundary in Marcus's stimuli is not clear, that for the present stimuli was due to inclusion of a labial release burst (from splash), which provided an additional stop manner cue (Repp, 1984).

Unlike the earlier studies, which used only temporal compression, the present experiment introduced both decreases and increases in acoustic segment duration. Although Marcus concluded from his results that the critical silent interval was invariant under changes of speaking rate, he failed to investigate the effects of decreases in simulated rate (i.e., increases in stimulus duration). According to the speaking-rate adjustment hypothesis, the perceptual boundary should shift to longer values of silence in that case, since no perceptual limit is encountered in that direction.

The question in Experiment 1 was, then, whether either "[s] duration" or "[l] duration," or both, have any effect on a short-silence [sl]-[spl] boundary.

Method

Stimuli. The utterances slash and splash were recorded by a female speaker, low-pass filtered at 9.6 kHz, and digitized at 20 kHz. To avoid strong stop manner cues in the [s] portion, the fricative noise of slash was used in all experimental stimuli. The remainder was taken from splash. This portion included an initial 10-ms release burst, which preceded the first glottal pulse of the [l] segment. The end of the [l] resonance was defined visually by a change in waveform shape coupled with an amplitude increase, and was confirmed by listening. The durations of the [s] noise and of the [l]

resonance were varied independently by either removing or duplicating a piece of the waveform. An appropriate piece was selected from the interior of each acoustic segment on the basis of overall and local envelope considerations, and all cuts were made at zero crossings. In the [s] noise (original duration: 142 ms), the piece removed or duplicated was 51 ms long and ended 36 ms before noise offset. In the [l] portion (original duration: 57 ms, or 14 pitch periods), it was 21 ms (5 pitch periods) long and began 28 ms (7 pitch periods) after [l] onset. Thus, the signal portions immediately adjacent to the closure interval were left undisturbed, so as to avoid changing spectral and amplitude envelope cues to stop manner (cf. Summerfield et al., 1981: Exp. 1). The ultimate durations were 91, 142, and 193 ms for the [s] noise, and 36, 57, and 78 ms for the [l] resonance. (Note that the changes are proportional and correspond to increases or decreases of about 36 percent.) The orthogonal combination of all [s] and [l] durations resulted in nine stimuli, for each of which silent closure duration was varied from 0 to 50 ms in 10-ms steps. The resulting 54 stimuli were recorded in 5 different randomizations with interstimulus intervals of 2 s.

Subjects and procedure. Seven paid volunteers and the author identified the stimuli as slash or splash, with stlash as an additional option. The tape was repeated once, so that 10 responses per subject were obtained for each stimulus. Presentation was over TDH-39 earphones at a comfortable intensity in a quiet room.

Results and Discussion

The results are shown in Table 1 in terms of category boundary locations, determined from the average labeling functions by linear interpolation. (Only three subjects gave any stlash responses, which were included with splash responses.) Repeated-measures analysis of variance was conducted on individual subjects' response percentages, averaging over silence durations. Increasing silence duration, of course, had the expected effect of increasing the percentage of splash responses; the labeling functions, which are not presented here for the sake of conciseness, were comparable in steepness to those obtained by Marcus (1978). As can be seen in Table 1, the amount of silence needed to hear a p (or t) increased as the duration of the [s] noise increased, $F(2,14) = 12.5$, $p < .001$, but decreased as the duration of the [l] resonance increased, $F(2,14) = 15.8$, $p < .001$. Both effects were highly consistent across subjects, approximately linear, and of similar magnitude. Their interaction was not significant, $F(4,28) = 1.1$.

Since increases and decreases in acoustic segment duration effected boundary shifts of nearly equal magnitude, it appears that the [sl]-[spl] boundary was not close to a lower limit. In fact, the boundary shifted to as little as 17 ms of silence in the "short [s], long [l]" condition, which is considerably shorter than any of Marcus's (1978) values. This suggests that Marcus's failure to find any boundary shifts was not due to the relatively short category boundary for his stimuli. Indeed, closer inspection of Table 1 reveals that Marcus's results are replicated by the present study: Due to the opposite and equally-sized effects of changes in [s] and [l] duration, simultaneous proportional compression or expansion of both acoustic segments had no effect on the [sl]-[spl] boundary. (Compare values in italics along the major diagonal in Table 1; $F(2,14) = 0.1$.) Thus, to the extent that the combined [s][l] duration conveyed anything about speaking rate, there was no effect of this variable in the present study.

Table 1

Results of Experiment 1: Average category boundary values (in ms of silence) as a function of [s] and [l] durations.

[l] duration (ms)	[s] duration (ms)			
	91	142	193	Mean
36	24.0	25.5	27.6	25.7
57	18.8	23.7	24.2	22.2
78	17.2	17.9	23.8	19.6
Mean	20.0	22.4	25.2	22.5

The observed effect of [s] duration on stop manner perception may be attributed to the "rate" of the speech preceding the silence, which really amounts to merely redescribing the results. An alternative explanation is in terms of a perceptual compensation reflecting listeners' tacit knowledge of the coarticulatory shortening of [s] frication preceding a stop closure. An independent effect of fricative noise duration was also found by Summerfield et al. (1981: Exp. 1); however, they tentatively attributed it to a psychoacoustic effect of this variable on the perceived silence duration. This hypothesis cannot be ruled out on the basis of the present data. However, the "coarticulation-compensation" hypothesis proposed should perhaps be favored in view of many related findings (see Repp, 1982).

The reversed effect of [l] duration was totally unexpected. Since [l] duration in natural speech does not seem to covary with the presence versus absence of a preceding [p], it is unlikely that [l] duration has any direct cue value for stop manner perception, in the way that [s] duration has. Rather than affecting stop manner perception directly, [l] duration may have its effect by altering the perceived relative duration of the [s] noise. (See Repp, Liberman, Eccardt, and Pesetsky, 1978, for a rather similar argument relating to the fricative-affricate contrast.) In other words, the [s] noise may "sound longer" before a short [l], and shorter before a long [l]. This explanation assumes that the intervening silence does not engage in such contrastive interactions with the surrounding signal portions; this assumption is supported by the absence of any effect of increases or decreases in both [s] and [l] duration.

Experiment 2

It is not yet clear why Summerfield et al. (1981) did find an effect of overall stimulus compression. One possibility is that their compression technique affected the amplitude envelopes of the signal surrounding the silence, thus introducing additional stop manner cues that shortened the amount of silence required to hear a p. However, since their technique was similar to

Marcus's, and in fact left about 10 ms of waveform on either side of the silence undisturbed, this possibility seems unlikely. Another possibility is suggested by the results of Experiment 1, however: The hypothesis just proposed to explain the effect of [l] duration predicts that the relational dependence of perceived [s] duration on the context following the silence should decrease with increasing temporal separation. Thus, at the longer silent intervals that characterized the Summerfield et al. stimuli, the effect of [s] duration may have been larger than the (presumably) opposite effect of the signal duration following the silence, thus leading to a net effect in the same direction as that of [s] duration alone.

It is also true, of course, that Summerfield et al. varied the duration of the whole stimulus, and not just of [s] and [l]. It was decided, therefore, to replicate their study using stimuli that had the category boundary at a comparably long silent interval (which was achieved by removing the labial release burst from the stimuli of Experiment 1 and by shifting the range of silent intervals employed). The main difference was that, in Experiment 2, the durations of [s], [l], and of the final [æʃ] portion were varied independently, so as to determine their separate effects on the slash-splash boundary.

Method

Stimuli. The 10-ms release burst was removed from the stimuli of Experiment 1. Two [s] and two [l] durations were employed, corresponding to the original and shortened versions of Experiment 1. In addition, the final [æʃ] portion was used both in its original version (477 ms) and shortened by 36 percent (304 ms). Shortening was achieved by deleting two separate pieces of waveform from the interior of the [æ] vowel and one piece from the interior of the [ʃ] noise, thereby reducing each of these two acoustic segments by the same proportional amount. Careful listening indicated no obvious disruptions of spectral continuity caused by the splices. The two [s] durations, two [l] durations, and two [æʃ] durations were combined to yield eight stimuli that were presented with six different silent intervals ranging from 50 to 100 ms in 10-ms steps. The resulting 48 stimuli were recorded in five randomizations with interstimulus intervals of 2 s.

Subjects and procedure. Ten paid volunteers listened to the tape twice, labeling each stimulus as slash or splash.[1] None of the subjects had participated in Experiment 1. The stlash response category was not included, since these responses generally occur only at short closure durations (cf. Repp, in press). Otherwise, the procedure was identical to that in Experiment 1.

Results and Discussion

The results are displayed in Table 2. The average labeling functions from which the boundaries were derived were less steep than in Experiment 1 but comparable to those obtained by Summerfield et al. (1981). The boundaries were located at somewhat longer silences than in the Summerfield et al. study, probably owing to procedural differences. It can be seen in Table 2 that the basic findings of Experiment 1 were replicated: The amount of silence needed to perceive splash increased as [s] duration increased, $F(1,9) = 42.3$, $p < .001$, and decreased as [l] duration increased, $F(1,9) = 20.5$, $p < .002$. As in Experiment 1, these effects were highly consistent and independent of each

other (F = 0.0 for their interaction). In contrast to Experiment 1, however, and in agreement with the predictions for Experiment 2, the effect of [s] duration was larger than the opposite effect of [l] duration, which supports the hypothesis that the latter effect is indirect and decreases with increasing temporal separation between [s] and [l], relative to the effect of [s] duration. In a separate comparison of the results for the two stimuli that differed by a uniform compression of 36 percent (values in italics in Table 2), a significant 6.5-ms boundary shift was observed, $F(1,9) = 15.2$, $p < .004$, which is comparable to the shifts found by Summerfield et al.

Table 2

Results of Experiment 2: Average category boundary values (in ms of silence) as a function of [s], [l], and [æʃ] durations.

[l] duration	[æʃ] duration	[s] duration (ms)		
		91	142	Mean
36	304	66.7	75.4	
	477	69.6	81.7	
	Mean	68.2	78.6	73.4
57	304	62.6	75.3	
	477	64.4	73.2	
	Mean	63.5	74.3	68.9
Mean		65.9	76.5	71.2

The effect of the duration of the [æʃ] portion was less consistent. There was a small but significant main effect, $F(1,9) = 6.2$, $p < .04$, as well as a significant interaction with [l] duration, $F(1,9) = 10.4$, $p < .02$. As can be seen in Table 2, the effect of [æʃ] duration was reversed with respect to that of [l] duration, longer [æʃ] durations leading to longer category boundaries, except in the condition where both [s] and [l] were long. While the reason for this interaction is not clear, the direction of the main effect suggests that, rather than influencing perceived [s] duration, the [æʃ] portion may have modified the perceived [l] duration, which then in turn influenced the perceived [s] duration. In other words, there may be a general contrastive interaction between adjacent energy-carrying acoustic segments of the speech signal with respect to their effective temporal features in phonetic perception. The effect of [æʃ] duration (but not that of [l] duration) is also consistent with a "contextual speech rate" explanation, but is too small to be of any significance. Clearly, the dominant effect is that of [s] duration.

General Discussion

The present results eliminate the apparent contradiction between the earlier results of Marcus (1978) and Summerfield et al. (1981), and they also rule out some of the interpretations advanced by these authors. They suggest that Marcus's failure to find a shift of the [sl]-[spl] boundary as a function of stimulus compression was due neither to a perceptual limit, nor to any insensitivity of the boundary to contextual influences. Rather, as Experiment 1 has shown, even boundaries at very short silences are highly sensitive to context and shift freely to both longer and shorter silences. (See also Repp, 1983: Exp. 4, for a shift to very short silences induced by a restricted stimulus range.) The absence of a net effect of stimulus compression or expansion when the silence duration is short seems to be due to the presence of two opposite effects, of [s] duration and [l] duration, respectively, which are equally strong and thus cancel each other out. Another way of expressing this result is that the [s]/[l] duration ratio remains constant at the phonetic boundary. On the other hand, Experiment 2 has shown that, when the silence durations are longer (as in the study by Summerfield et al.), the [s] duration effect is larger than the [l] duration effect, so an effect of overall compression is obtained. This overall effect does not seem to reflect an adjustment to perceived global speaking rate but may be due to [s] duration alone, assuming that [s] duration is perceived relative to the context following the silence. As the temporal separation increases, the influence of this context on perceived [s] duration decreases in importance.

The effect of [s] duration is interpreted here as a perceptual compensation for the known reduction in fricative noise duration when it precedes a stop consonant closure. Thus it is considered a purely phonetic effect, deriving from listeners' tacit knowledge of speech patterns (Repp, 1982). This hypothesis predicts that no such effect should be obtained in analogous nonspeech stimuli--a prediction that obviously should be tested. In a more speculative vein, the reversed effect of [l] duration is attributed to some form of perceptual contrast among temporal stimulus properties. It is not clear at what level in perception this contrast might arise, but experiments with nonspeech stimuli should also prove revealing in that regard.

While the present results disconfirm the hypothesis that the [sl]-[spl] boundary shifts as a function of global contextual speech rate, the data are compatible with the assumption that listeners compute a variable running estimate of speaking rate on the basis of local temporal properties of the speech signal. In fact, this alternative hypothesis allows for contrastive interactions among adjacent segments whose relative durations deviate from the ratios commonly encountered in natural speech. Accordingly, the effect of [s] duration may be attributed to the listener's estimate of the local speaking rate at that time, based on [s] duration relative to the following context. While this account provides an alternative to the hypothesis of perceptual compensation for [s]-stop coarticulation, the latter hypothesis is to be preferred because speaking rate is not a quantity that varies from segment to segment in speech production; hence, to postulate a corresponding, continuously varying perceptual dimension is of questionable explanatory value. Moreover, as Miller, Aibel, and Green (1984) have recently shown, perceptual effects of local temporal stimulus properties are independent of subjects' estimates of (global) rate of articulation. The only serious alternative to the coarticulation-compensation account, therefore, is a purely psychoacoustic explanation based on auditory temporal contrast, which needs to be tested directly in future experiments.

References

Bastian, J., Eimas, P. D., & Liberman, A. M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. Journal of the Acoustical Society of America, 33, 842. (Abstract)

Fitch, H. L., Halwes, T., Erickson, D. M., & Liberman, A. M. (1980). Perceptual equivalence of two acoustic cues for stop-consonant manner. Perception & Psychophysics, 27, 343-350.

Haggard, M. (1973). Abbreviation of consonants in English pre- and post-vocalic clusters. Journal of Phonetics, 1, 9-24.

Marcus, S. M. (1978). Distinguishing 'slit' and 'split'--an invariant timing cue in speech perception. Perception & Psychophysics, 23, 58-60.

Miller, J. L., Aibel, I. L., & Green, K. (1984). On the nature of rate-dependent processing during phonetic perception. Perception & Psychophysics, 35, 5-15.

Morse, P. A., Eilers, R. E., & Gavin, W. J. (1982). The perception of the sound of silence in early infancy. Child Development, 53, 189-195.

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. Perception & Psychophysics, 32, 141-152.

Repp, B. H. (1979). Influence of vocalic environment on perception of silence in speech. Haskins Laboratories Status Report on Speech Research, SR-57, 267-290.

Repp, B. H. (1980). A range-frequency effect on perception of silence in speech. Haskins Laboratories Status Report on Speech Research, SR-61, 151-165.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. Psychological Bulletin, 92, 81-110.

Repp, B. H. (1983). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. Speech Communication, 2, 341-362.

Repp, B. H. (1984). The role of release bursts in the perception of stops after [s]. Journal of the Acoustical Society of America, 75, 1219-1230.

Repp, B. H. (in press). Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. Language and Speech.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. Journal of Experimental Psychology: Human Perception and Performance, 4, 621-637.

Schwartz, M. F. (1969). Influence of vowel environment upon the duration of /s/ and /s/. Journal of the Acoustical Society of America, 46, 480.

Schwartz, M. F. (1970). Duration of /s/ in /s̆/-plosive blends. Journal of the Acoustical Society of America, 47, 1143-1144.

Summerfield, Q., Bailey, P. J., Seton, J., & Dorman, M. F. (1981). Fricative envelope parameters and silent intervals in distinguishing 'slit' and 'split'. Phonetica, 38, 181-192.

Footnote

[1]Subjects' comments after a brief preview of the tape revealed that most stimuli were initially perceived as splash. All subjects were consequently encouraged to try to hear more instances of slash, and to classify ambiguous stimuli as belonging to this category. No subject had any difficulty carrying out these instructions, which were probably unnecessary, since phonetic boundaries based on closure silence duration are rather sensitive to the range of

closure durations employed in a test (Repp, 1980). Most likely, the listeners in Experiment 2 rapidly adjusted their

THE PHYSICS OF CONTROLLED COLLISIONS: A REVERIE ABOUT LOCOMOTION*

Peter N. Kugler,† M. T. Turvey,†† Claudia Carello,††† and Robert Shaw††††

> "No fact of behavior, it seems to me, betrays the weakness of the old concept of visual stimuli so much as the achieving of contact without collision--for example, the fact that a bee can land on a flower without blundering into it. The reason can only be that centrifugal flow of the structure of the bee's optic array specifies locomotion and controls the flow of locomotor responses"
>
> "But to understand, to be able to explain and predict, entails the knowing of laws. It is our own fault if we do not know the laws"
>
> (From Gibson's autobiography in E. Reed & R. Jones (Eds.), Reasons for realism: Selected essays of James J. Gibson, Hillsdale, NJ: Erlbaum, 1982, pp. 14 and 15, respectively.)

Introduction

Imagine the following scenario. It is late in the afternoon and since early morning you have been mulling over a long-term concern of Gibson's (1950, 1960, 1961, 1966, 1979), namely, the optical structure ambient to an animal that is generated by the layout of surfaces and by the animal's movements (both the movements of its limbs relative to its body and the movements of its body, as a unit, relative to the surface layout). You are taken by the subtlety of Gibson's point that this optical structure resembles neither the surface layout nor the movements but it is specific to them because it is nomically (lawfully) dependent on them. And you are impressed by Gibson's insistence that these dependencies between properties of the animal-environment relation and properties of the ambient light are instances of laws, indigenous to the ecological scale (the scale of animals and their environments), that make possible the control of activity.

---

*To appear in W. H. Warren, Jr. & R. E. Shaw (Eds.), Persistence and change: Proceedings of the First International Conference on Event Perception. Hillsdale, NJ: Erlbaum, in press.

†Crump Institute for Medical Engineering, University of California, Los Angeles.

††Also University of Connecticut.

†††State University of New York at Binghamton.

††††University of Connecticut.

Acknowledgment. This work was supported in part by NICHD Grant HD-01994. The authors wish to thank A. S. Iberall for his comments on parts of this chapter.

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

Your thoughts return repeatedly to locomotion, Gibson's favorite example, and to his characterization of locomotion as a matter of controlled encounters (Gibson, 1979) with the substantial surfaces that comprise the objects and places of the animal's niche. In the course of locomoting, an animal's surfaces may contact surfaces of the environment. These contacts are selective and they vary in intensity. There are hard contacts (as in predatory attacks), medium contacts (as in diving into water), soft contacts (as in alighting on a branch) and non-contacts (as in steering between trees). It seems to you that it might prove helpful to know what happens to bodies, in general, when they collide. And to this purpose, you direct your reading to the physics of collisions (summarized in the Appendix).

Your attention begins to wander. Looking out the window you see a bird in flight (Figure 1). You admire its ability to adjust its flight to the surroundings. Your thoughts meander--"laws," "controlled collisions," "a physics of the ecological scale." You fall asleep and dream...

The Reverie

You are a physicist investigating a type of visible particle whose identity is unknown to you. Particles of this type range in mass from .001 kg to 10,000 kg. You watch the trajectory of a token particle through a non-uniform, three-dimensional surround as depicted in Figure 2. In some regions of the surround, matter or energy is more concentrated than in other regions. The particle sometimes moves between the particularly dense regions and sometimes it contacts them. The particle's speed is not uniform. There are obvious decelerations and accelerations prior to contact, but these are not uniform either. Sometimes contact is preceded by a marked deceleration so that the contact is gentle--very little momentum is exchanged. Sometimes the deceleration prior to contact is hardly noticeable or there is an obvious acceleration so that the contact is violent or hard--a great deal of momentum is transferred to the contacted region. And sometimes the deceleration is in an intermediate range, such that the contact is neither gentle nor especially violent.

Not all of the particularly dense regions of the surround are stationary. Some regions move just like the particle. Other regions move, but without the variations in accelerations that characterize the particle. Basically, the particle's trajectory with respect to the moving parts of the surround is not different from its trajectory with respect to the stationary parts: there is a steering among moving regions and contact--ranging from soft to hard--with moving regions.

Repeated observation of the particle's behavior with respect to the surround leads you to certain tentative conclusions as to its nature.

Conclusion 1: In tracking the particle's behavior, you monitor the mechanical quantity of momentum. The rate of change of momentum identifies a force or interaction between particle and surround. Usually momentum and its first derivative prove sufficient for the purpose of describing a given particle's trajectory. For the behavior of this particle, however, it seems that there is another mechanical quantity that is much more relevant: the second derivative of momentum or the rate of change of force. Characteristically, as the particle approaches a region of the surround, it exhibits a systematic sequence of accelerative changes. You wish to give this mechanical quantity a

Figure 1. Observing controlled collisions.

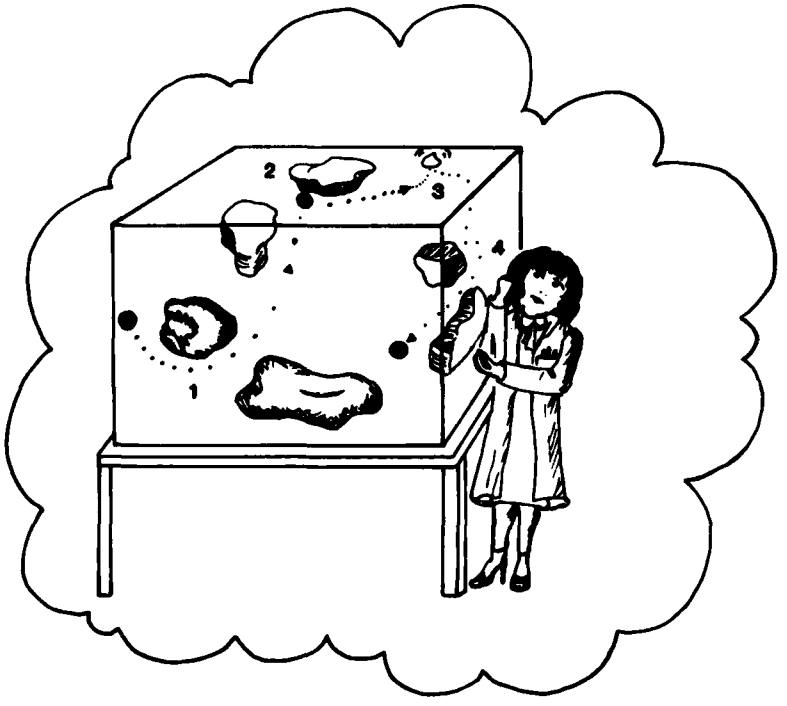


Figure 2. As the particle moves through a non-uniform surround, it sometimes steers between dense regions (1 and 4) sometimes contacts them gently (2) or violently (3), and does not maintain a uniform speed.

name. "Jitter" comes to mind, but for obvious reasons you are attracted to "control" and you make note of the control quantity's relation to the more familiar mechanical quantities of momentum, impulse, and force (Table 1).

Table 1

QUANTITY	SYMBOL	COMPOSITION	DIMENSIONS
MOMENTUM	P	MV	MLT^{-1}
IMPULSE	I	ΔMV	MLT^{-1}
FORCE	F	$\Delta MV/T$	MLT^{-2}
CONTROL	C	$\Delta MV/T^2$	MLT^{-3}

[Where M = mass, V = velocity, T = time, Δ = change, L = length.]

The control of a collision (read in the same sense that one would read "the momentum of a collision" or "the force of a collision") is, therefore, measurable. It would be given by the integration of C within the spatial and/or temporal limits of the collision, assuming that they can be reasonably approximated. Because of the fact that the mechanical quantity of control is a natural extension of the mechanical quantity of force, you are willing to speculate that there is a (scalar) quantity that relates to control in the manner that potential (a term referring to the concentration or distribution of a conserved quantity such as energy) relates to force. Ordinary language usage suggests the term coordination for this quantity. The suggestion is fortunate: Both "potential" and "coordination" are configurational notions. You are tantalized by this idea that the conceptions of control and coordination may be interpreted as mechanical quantities that are as principled in their relation to one another as are force and potential.

Conclusion 2: It is evident that while proximity to things in the surround is a determinant of the forces forming the particle's trajectory, it is neither the sole determinant nor the most significant. Conventional particle trajectories are shaped by interactions with regions--usually other particles--that attract or repel a particle to varying degrees depending on the particle's distance from them. A force that is a function only of distance is termed "conservative." The forces affecting the trajectory of your particle seem to depend on time (the time to contact) and, perhaps, velocity (the velocity prior to contact). They are non-conservative forces. You guess that these forces--which entail a dissipation rather than a conservation of energy--originate in the particle rather than in the surround. There is something special about this particle; it seems to have (on board) a replenishable source of potential energy that it can deploy.

Conclusion 3: The number of soft, medium, hard, and non- collisions exhibited by your particle during a period of observation is very large. Given so many interactions, you think it worthwhile to adopt a statistical mechanical orientation toward the particle's behavior. It seems particularly promising to inquire about the distribution function that characterizes the

many interactions of the particle and surround. In the tradition of Boltzman, Maxwell, and Zipf you look to the distribution function as a way of appreciating the constraints--the quantities that must be conserved--on the interactions of particle-like entities. Relatedly, you see the usefulness of the distribution function for classifying particles. Types of interactions will be broadly distinguished by the quantities conserved over interactions; these differences in conservations will show up as differences in distribution functions given that a distribution function is completely determined by the operative conservations.

In the construction of a distribution function one asks, roughly, how many particles (in any arbitrarily chosen volume) will possess a particular value of a particular quantity. Boltzman and Maxwell focussed on gases and the property of velocity. Over the very many interactions of n particles of a gas, the conservations of total mass (nmv_0), momentum (nmv_1) and vis viva (nmv_2, or twice the kinetic energy) determine that the particles will tend to move at one particular speed, more or less. Collectively, the conservations select ("prefer") a distance between collisions (mean free path) and a time between collisions (mean relaxation time). The mean and variance (the "more or less") of the velocity reflect the concentration of the conserved quantities. The mean and the variance of the velocity prove to be characteristics of a gas, and both are affected by its temperature.

Thinking about your particle in comparison to a gas particle, you are of the opinion that the contrast between the two is most sharply drawn with respect to momentum change in relation to velocity. Impulses of gas particles are of maximal frequency when the velocity of the particles is zero, that is, at the moment of impact. At any other moment impulse is nonexistent. Statistically, your particle could be assigned a mean free path and a mean relaxation time but, importantly, across the full range of velocities that it exhibits, impulses can be observed. Unlike the case with gas particles, there is the velocity at which the frequency of impulses is concentrated.

You imagine a distribution function defined over three coordinates: number of particles, velocity, and number of impulses. For a typical gas and for particles of the type you are studying, the distribution functions differ significantly. The peaking of impulse frequency at zero velocity that reflects the conservations governing the gas will not be found in the distribution function of your particle type. What does the absence of a peak (the fact that impulse is uniformly distributed over velocity) mean? The distribution function for your type of particle must be the way it is because of the conservations that are operative. This is true by definition. However, the conservations governing your particle's behavior cannot be the typical velocity-linked conservations of mass, momentum, and energy. Governing your particle's behavior are conservations that are velocity indifferent.

Conclusion 4: Although you are unable for the present to say much about the selectivity of the trajectory--the fact that some regions function as attractors and some as repellers--it is clear to you that the particle's trajectory minimizes the momentum transferred to the particle from the surround. To what sort of principle is the particle subject that demands no momentum bumps? If the particle's interior was complex and if its persistence depended on maintaining that interior, then keeping the level of momentum absorption below some critical value would clearly be important--large transfers of momentum could fracture the particle (see Appendix). At the lev-

el of the particle this principle reads: Move so as to conserve a smooth unitary process ('smooth' meaning no sudden energy or momentum bumps--excessive energy or momentum exchanges--and 'unitary' meaning that the characteristic form and function of the particle is preserved). As a physicist, however, you might be uncomfortable with a conservation that is (1) defined at the level of the individual particle and (2) not identified with a quantity. The traditional conservations of mass, energy and momentum are in reference to measurable physical quantities exhibited by the particle. Further, the invariant nature of these quantities is not defined at the level of the individual particle, but minimally at the level of a pair of interacting particles. For example, with regard to momentum conservation, the momentum of each of two individual particles may change with a collision but the summed momentum of the two particles after collision equals the summed momentum of the two particles before collision.

Your discomfort with the notion of a conservation of a smooth, unitary process might be alleviated (but not eliminated) by the observation that some of the so-called quantum numbers conserved in the collisions of sub-atomic particles denote a qualitative property--the class of the particle--that is invariant at the level of the individual particle. You note how well leptons (approximately eight particles that do not take part in "strong" interactions) conserve their class membership; accelerating a lepton such as the positron to the point where its mass is equal to that of a proton (a member of the baryon class of particles that do take part in "strong" interactions) does not result in a metamorphosis. Nevertheless, you would be happier with a more traditional orientation to conservation, given the size of the particle you are studying. You suppose that your particle might be a member of a class. Is there a conserved quantity defined at the level of the class? For example, over the many trajectories of the many members of this class, perhaps the number of members is conserved.[1] If a quantity such as the latter had to remain constant, then the minimization of momentum transfer from surround to particle (and hence the conservation of a smooth, unitary process) would be rationalized.

Conclusion 5: You recognize that a circumstance, such as the one you are studying, in which forces are shaped to achieve one trajectory and to prevent others, usually defines a machine. Somehow a machine conception must be brought to bear on your understanding of the particle. Because a machine is a way of harnessing mechanical forces to do work in determinate directions, a machine can be properly termed a constraint--a restriction on the laws of motion. Very often a machine is constructed with hard, resistant pieces linked by hard, resistant chains. Is your particle a hard-molded machine like this? What makes you dubious is that a hard-molded machine is not very flexible and the particle's trajectory indicates that the shaping of force to achieve gentle, medium, and violent collisions, or to avoid collisions, is flexible. The rate of change in the rate of change of the particle's momentum (i.e., the control) varies from region to region of the surround. The unavoidable conclusion is that the forces are harnessed by a constraint that cannot be hard-molded. To draw the comparison, you might say that the constraint on the non-conservative forces centered in the particle is "soft" rather than "hard" and that the appropriate machine conception is soft-molded rather than hard-molded.

Conclusion 6: Obliged to avoid action at a distance you make the assumption that the soft constraint on the particle-based forces is a field. This field is ambient to the particle. Is it a field associated with a force, a quantum field? Of the four fundamental forces, only the gravitational and electromagnetic forces apply, given the magnitude of the particle. The electromagnetic field would seem to be a better bet than the gravitational, but neither is particularly appealing because you are convinced that if the soft constraint is a field, it cannot be a field associated with a force. It may well be caused by electromagnetic phenomena but it is qualitatively different from them.[2] Your conclusion follows in part from certain distinctions drawn by Pattee (1972, 1977). Forces and constraints are not things of the same type, even though constraints--like all other aspects of nature--are built from the four fundamental forces. To begin with, the forces are not embodied in anything particular and they apply to everything within the range to which they apply (gravity, for instance, applies everywhere). A constraint, however, has a particular embodiment and applies to a particular thing. Further, whereas the important feature of a force, its magnitude, is dependent on rate (the derivative of a variable or variables with respect to time), the important feature of a constraint, its selectivity (resulting in one directed motion rather than others), is not dependent on rate.

Conclusion 7: It is a small step from the preceding conclusion to the conclusion that if the field in question is not a force field, then the fundamental dimensions from which its relevant variables are constructed cannot include mass (M). That is, the field must be kinematic--of fundamental dimensions length (L) and time (T)--or geometric--of fundamental dimension L, but it cannot be kinetic--of fundamental dimensions, M, L, and T.[3] As you have already noted, this field must constrain the dissipative forces focused in the particle so as to keep to a minimum the momentum transferred to the particle from the surround. You puzzle over this requirement. Doesn't it mean that the field in question must be structured by the kinetics of the surround and the kinetics of the particle? If the field did not faithfully reflect these two kinetic domains, then there would be no lawful basis for relating forces originating in the surround to forces originating in the particle, and the exchange of momentum could not then be regulated. You suppose, therefore, that the field in question has this capability and inquire what this tells you about the general properties of the field.

To bring things into focus, you assume (i) the particle to be in motion at a constant velocity in one direction and (ii) an absence of motion in the surround. Normally you would represent this by a velocity vector originating in the particle and pointing in the direction of travel. However, you find it convenient to think of the field hydrodynamically--as a fluid flowing relative to the particle. So instead of assigning a velocity vector to the particle (because you regard it as the origin), you assign a velocity vector (the negative of the particle's velocity) to each point in the field, where each field point can be anchored to a surround point.

This vector flow field viewed strictly as a kinematic field is always at equilibrium; subsequent to a disturbance there is no tendency on the part of the field to restore the structure it had prior to the disturbance. Further, from the perspective of the flow field, a disturbance is reversible in that any disturbance and its reverse are energetically equal. This reversible, equilibrium character of the flow field is because the flow field is not paying the energy cost, so to speak, of its changes. That bill is being paid by the kinetic field--the particle--to which the flow field is coupled: Only

changes in energy flux can give rise to changes in flow, and the changes in energy flux in this case are bound to the particle's on-board energy reservoirs or potentials.

The reversibility of the flow field appears to be of paramount importance. If the flow field were not reversible, if it carried potentials that "wound-up" the trajectories, then the flow field would itself determine some of the properties it exhibits. A reversible field, on the other hand, meets the criteria of linearity--superposition and proportionality--and can, therefore, faithfully map the kinetics that give rise to it. You feel that there may well be a very general principle here: The availability of a reversible field is a prerequisite for the kind of controlled collisions that your particle exhibits with respect to its surround.

What properties arise in the flow field caused by the particle's motion relative to the surround? A coarse analysis reveals the following: kinematic properties, consisting of (i) transformations defined over the entire flow field--such as outflow from a point and inflow to a point--and (ii) the inverse of the rate of dilation of a topologically closed region of the field; and (iii) geometric properties, viz., singularities, such as foci of outflow and inflow.[4] Global transformations ((i) above) are specific to the displacement of the particle as a unit relative to the surroundings (moving forward or backward); the inverse of the rate of dilation (ii)--a property you recall reading about in the astronomer Hoyle's science-fiction novel The Black Cloud (1957)--is specific to the time at which the particle will contact a region on its path while the first derivative of this property, which is seen to be a dimensionless quantity, is specific to the deceleration of the particle with respect to the approach region.[5] The foci of flow (iii) will be specific to the regions, or to the gaps between them, toward which the particle is moving; that is, the foci are specific to the direction of the particle's trajectory.

It is obvious to you that under normal circumstances, the style and/or rate of transformation will not be uniform throughout the entire kinematic field; rather, there will be discontinuities caused by region boundaries that will identify more precisely the relationship between the moving particle and a particular layout of dense regions (depots of mass). For example, within the global outflow "local" properties will be revealed, such as: (i) a gain of structure inside a closed contour in the field specifies an opening in a dense region through which the particle could travel, (ii) a loss of structure outside a closed contour in the field specifies an obstacle to the particle's current trajectory.[6]

Clearly, motion of the particle gives rise to properties that do not exist when the particle is immobile. The properties identified above, both kinematic and geometric, are annihilated when the temporal dimension goes to zero and the ambient kinematic field is reduced to an ambient geometric field. For example, "streaming" engendered by the particle's motion condenses out geometric, rate-independent points, the singularities, that are not identified by a geometric field analysis. A geometric field analysis at any instant of time would not contain the singularities.

Conclusion 8: You are drawn to the fact that your cursory examination of the properties of the kinematic field (caused by the displacement of the particle relative to the surround) revealed a dimensionless number: The first derivative of a kinematic field property specifying time-to-contact. What

intrigues you is the possibility of an analogy between the dimensionless quantities of the kinematic field (assuming that there are more to be discovered) and the dimensionless quantities that order a kinetic field, such as a hydrodynamic field.

The transition from one state to a qualitatively distinct state of a physical system usually indexes a critical change in the relation between two competing processes. Your favorite example is the transition from laminar flow to turbulence, which occurs when the processes (viscous, dissipative, irreversible) that resist fluid motion cannot, in their current organization, balance the processes (inertial, conservative, reversible) that sustain fluid motion. The dimensionless Reynolds number gives an index of the competition between inertial (etc.) and viscous (etc.) processes. High inertial forces favor turbulence, with the pronounced internal shearing that that implies. High viscous forces prohibit sustained turbulence by damping motions that lead to discontinuity (e.g., eddies) and thus ensure laminar flow. The inertial processes are governed by Newton's law of inertia and the viscous processes are governed by the law for shear stress of a Newtonian fluid. The Reynolds number, therefore, might be described as indexing the relation between the two laws. On either side of the critical value of the Reynolds number the two laws are mutually cooperative, whereas at a critical value one of the two laws dominates the other (that is, a competition occurs).

You are aware that, as a general rule, any major dimensionless number used in physics can be derived directly from the laws known to apply to the phenomenon to which the number refers (Schuring, 1977). A dimensionless number is often referred to as a Pi number (Buckingham, 1914) and when it is derivable from one or more laws, it is termed a principal Pi number (Schuring, 1977). The important thing you note here is the linkage between physical states of affairs that principal Pi numbers index and the facts of critical values and behavioral modes (or natural categories). As you see it, the shift in balance between two (or more) laws governing a phenomenon from situations in which they cooperate to situations in which one law alone is responsible can produce categorically distinct states. The transition from cooperation to competition between governing laws is tantamount to a natural boundary-making device: behavioral modes are created, critical values of one or more variables are defined.

In sum, the critical values of dimensionless quantities in the kinetic cases mark off distinct physical states. It does not seem likely to you that dimensionless quantities will play this role in the kinematic field of constraint because of the absence of forces--by definition--in the kinematic field. But you cannot be too sure, one way or the other. For the present, however, it seems prudent to emphasize the specificational rather than the physical nature of the kinematic field. This emphasis raises the question: Do dimensionless quantities in the kinematic field mark off--at critical values--distinct specificational states?

A soft collision with no momentum exchange between the particle of interest and a nonmoving, dense region on its path requires that the particle decelerate. A deceleration is adequate if and only if the distance it will take the particle to stop with that deceleration is less than or equal to the particle's current distance from the region of upcoming contact. Your calculations show that for any particle of the type you are studying a deceleration is adequate if and only if

$$\text{Pi(contact)} = \frac{d\tau(t)}{dt} \geq -0.5$$

where $\tau(t)$ is the time-to-contact variable of the kinematic field.[7] You state this result as follows: When less than -0.5, the dimensionless quantity, Pi (contact), specifies that the particle will experience a momentum bump if present conditions persist; when equal to or greater than -0.5, Pi (contact) specifies that the particle's contact with the upcoming region will involve no momentum exchange if present conditions persist.

You are encouraged by the results of your analysis. It does seem that critical values of dimensionless quantities in the kinematic field distinguish between qualitatively distinct specificational states. And it seems to you that the analogy should be pursued further. For example, you might ask: What kinds of laws go into the construction of pi numbers applicable to the kinematic field?

Conclusion 9: Because the kinematic field ambient to the particle constrains its trajectory, you infer that the field and the particle must be coupled. This coupling is obviously "soft" rather than "hard." The question to which you now turn is: What must be required of the particle and of this soft-coupling if the particle is to be constrainable in a way that makes its collisions controllable? What must be true of the particle so that it can be reliably constrained by the kinematic field?

It appears to you that there are two important and very general conditions on the coupling. One condition is that the coupling be linear. What would have to be true of the particle's interior in order to guarantee a linear coupling? The interior of the particle could be in either a reversible or irreversible steady state. If it were reversible, the distribution of conserved quantities would be (nearly) uniform and the interior would be (approximately) at equilibrium. This means that there would be no problem of 'connectivity': A disturbance felt by any region of the interior could be transported to any other region of the interior, however remote. On the other hand, if the interior's steady state was irreversible, then there would be marked and persistent source-sink gradients. As a consequence, a disturbance felt in one part of the interior may not be transported to other parts. Conservations are not carried up gradients and, conventionally, it is through the transport of conserved quantities that one part of a physical system "informs" another part about what it is doing. A loss of connectivity among the regions that accompanies irreversible steady states means that the overall effects of the kinematic field on the particle's interior--however those effects are realized--could be discontinuous and equivocal. In short, it seems to you that if the steady state of the interior were irreversible and far from equilibrium, then there would not be a constant scale for laws relating properties of the kinematic field to force trajectories of the particle. You are led to assume, therefore, that a linear coupling, which would be both flexible and precise, requires a reversible, close-to-equilibrium steady state. This is tantamount to assuming that the state space of the particle's force trajectories are quasiergodic (that is, no strong preferences or dislikes): The particle should not be biased in a way that undercuts the specifying capability of the kinematic field.

The other condition on the coupling is that the criterial "smooth and unitary process" be upheld. This condition would be met only if the coupling involves very little energy (relative to the energy stored and dissipated by the particle). A coupling achieved at high energy expense might take too long (there would be steep external gradients) or it might involve a large momentum exchange and irreversible processes (marked by stress and shock waves). You conclude that there must be an energetically cheap translational gate effecting the coupling of the particle to the kinematic field.[8] Or, said differently, you conclude that the kinematic field is the spatio-temporal structure of a low energy field. Your best hunch is that this low energy field is the electromagnetic field modulated by the absorption/emission properties of the surround.

Conclusion 10: Some of your observations of the particle's trajectories are especially puzzling. Two of them are depicted in Figures 3 and 4. In one observation (Figure 3), you noted that your particle mimicked the trajectory of another particle of like kind. The two trajectories were, for a time, coupled. This coupling of trajectories did not depend on the distance between the particles. Sometimes you witnessed the coupling when the particles were very close (Figure 3a). At other times you saw the coupling when the particles were separated by a substantial distance (Figure 3b).

In the other observation (Figure 4) you noted that your particle's trajectory would follow, without contact, the border of a dense region in the surround. Here it seemed that there was another temporary coupling--between the form of the particle's trajectory and the form of a region.

Why do you find these observations especially puzzling? It is because, as a physicist, you are committed to explaining any coupling (coordination or cooperativity) of one thing with another through conservation principles, and it is not immediately obvious to you what the principles are that apply to the two couplings depicted in Figures 3 and 4. If you had observed two, more conventional particles coupled in interaction, then you would have said that (1) some quantity was exchanged between the particles--at the very least momentum and energy; and (2) the coupling was an instance of coordination or cooperativity because the exchange of quantities between the particles is constrained by the requirement that these quantities be conserved over the pair of particles. You would explain the loss of degrees of freedom that marks an interaction between particles by an appeal to conservational invariants.

You feel, therefore, that you have no option but to identify the conservations that account for the coupling phenomena depicted in Figures 3 and 4. Because the "mimicking" phenomenon is indifferent to particle separation, you believe that the conservations in question are unlikely to be energy or momentum related. Conventionally, couplings based on energy exchange depend on the distance between the particles (i.e., the inverse square law).

After a good deal of deliberation and hesitation you suggest the following: One of the conservations accounting for phenomena of the type depicted in Figures 3 and 4 must be conservation of topological form. (You believe that this conservation is integral to these instances of cooperativity but recognize that this conservation alone cannot account for the loss of degrees of freedom.) Your use of topological form is intuitive rather than technical. You mean, most generally, adjacencies and successitivies--that is, neighborhoods in space and time. And you mean, more particularly, properties

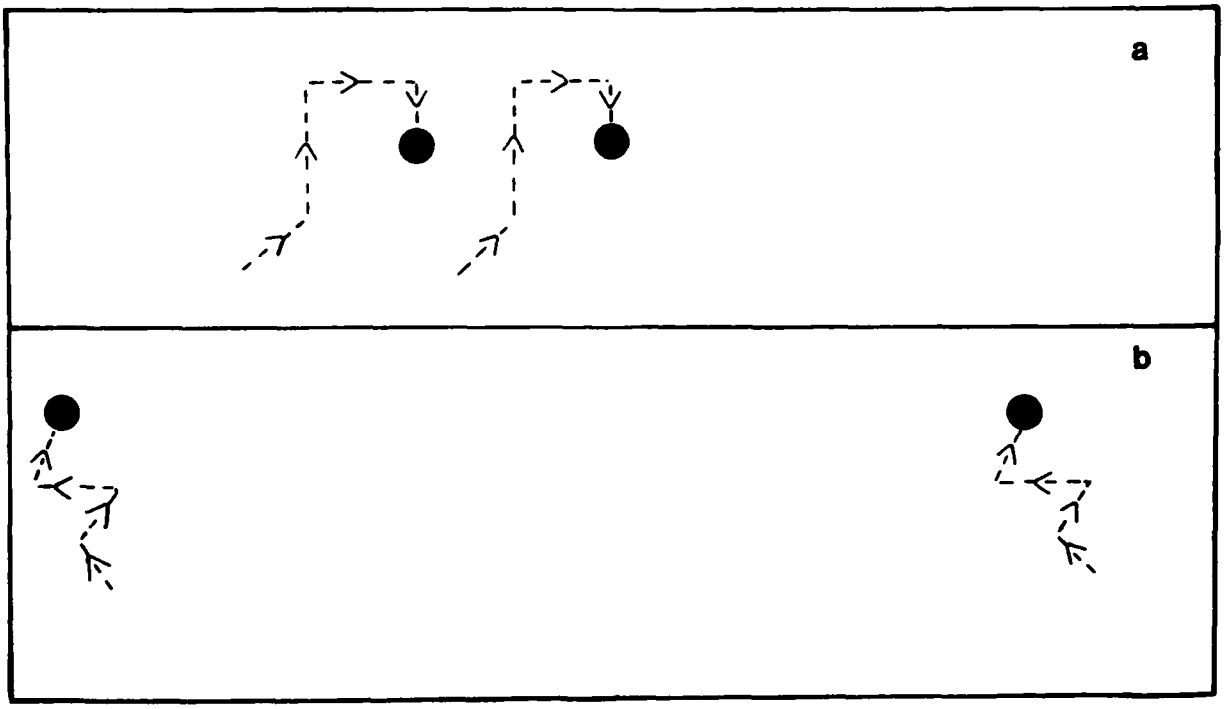


Figure 3. The particle mimicks the trajectory of another at near (a) and far (b) distances.

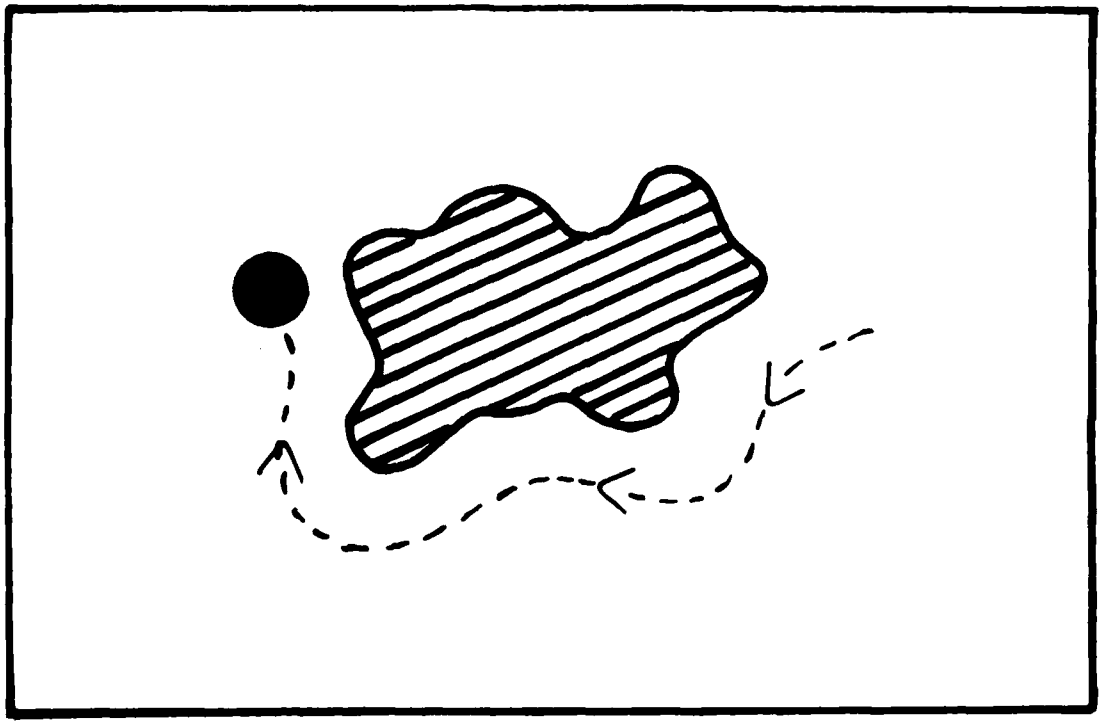


Figure 4. The particle's trajectory follows the border of a dense region of the surround without contacting it.

of the kind captured in contrasts such as inner/outer, sooner/later, lower/higher, closer/further, slower/faster, larger/smaller, and so on. Further, your use of conservation is intended to mean that from one "slice" to another, of the kinematic field that couples the particle to the surround, the topological form is constant. This conservation of adjacencies and successivities from a location proximate to the source to a location distal to the source is made possible by the reversible, equilibrium, low-energy nature of the kinematic field. Identifying the two particles in Figure 3 as kinetic fields, it is clear that the adjacencies and successivities arising from one kinetic field are perfectly conserved over the distance that separates the two kinetic fields. The proof is in the adjacencies and successivities arising from the second kinetic field (your particle)--they duplicate those arising from the first.

Conclusion 11: A better stab can now be made at the machine conception befitting the constraining of the forces that determine the particle's trajectory. You have come to the understanding that whatever the machine conception, it cannot apply just to the particle; rather, it must apply minimally to both the particle and to the kinematic field that is lawfully generated by the surround and the particle's displacement relative to it. It is very obviously true that the particle and the kinematic field are distinguishable. They clearly are different materially and, further, the particle, as a source of forces, is a kinetic field. Given that they are so different, you are puzzled by the principle that relates them as a single machine.

Now you are set to thinking: What, after all, is a machine? Turning to examples of hard-molded machines you are struck by the fact that they are always closed kinematic chains, where a chain consists of kinematic pairs of elements, for example, shaft and bearing, bolt and nut, etc. Each element in a pair, because of its resistant material qualities and its form, envelops and constrains the other so that all motions except those desired in the mechanism are prevented. There is kinematic closure. You can appreciate why a thoughtful student[9] of hard-molded machines might say that a machine consists solely of elements that correspond, pair wise, reciprocally. Kinematic closure is the central principle governing the construction of hard-molded machines.

Two other features of hard-molded machines capture your attention. First, in a closed pair of elements the roles of "fixed" and "movable" can be exchanged (for example, the nut can rotate and translate relative to the fixed bolt or the bolt can rotate and translate relative to the fixed nut). This inversion of roles causes no change in the motion belonging to the pair as you show in a sketch (Figure 5). In both of the situations shown in your sketch the separation between the nut and the head of the bolt is decreasing. Second, although it is common for a pair of elements to be completely closed in terms of bodily envelopment, it is not necessary. The closure that prevents certain motions from occurring can be achieved without material structures; you note, for example, how vertical downward closing forces keep the wheels of a train in contact with the rails.

It occurs to you that this invariant characteristic of hard-molded machines--reciprocally constraining, kinematic pairs--may well be an invariant characteristic of all machines, including the soft-molded machine you are trying to understand. Are the paired elements of this machine, the particle and the field ambient to the particle, kinematically closed? If there is a generalizable principle of kinematic closure, as you suppose, then the parti-

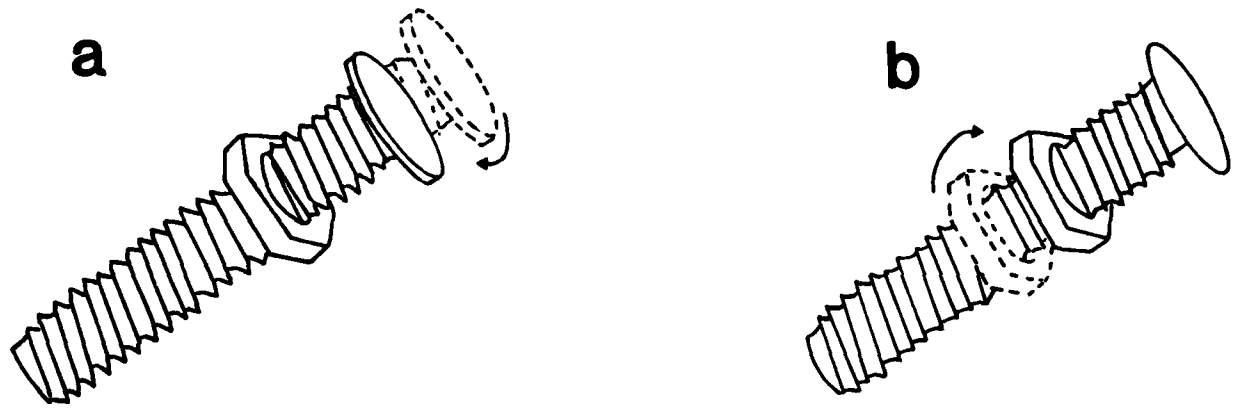


Figure 5. An example of a hard-molded machine. The distance between the nut and the head of the bolt can be decreased either by turning the bolt relative to the fixed nut as in (a) or turning the nut relative to the fixed bolt as in (b).

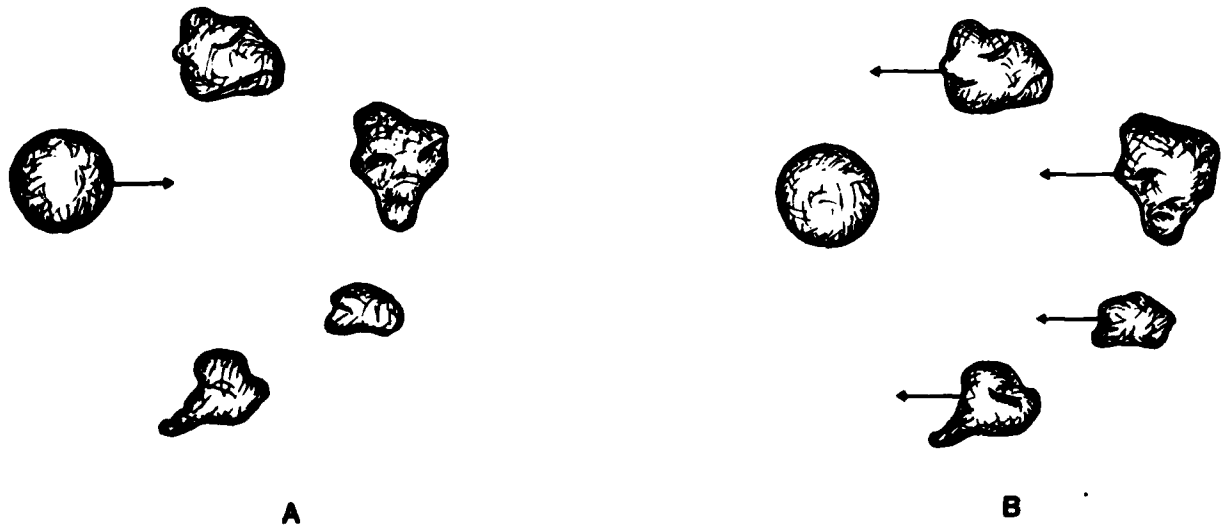


Figure 6. An example of a kinematically closed soft-molded machine. The distance between the particle and the surround can be decreased either by moving the particle relative to the fixed surround as in A or moving the surround relative to the fixed particle as in B.

cle and the ambient field should pass the inversion test: For example, fixing the entire surround and moving the particle in one direction should have the same consequence as fixing the particle and moving the entire surround in the opposite direction. In Figure 6, situation A should be indistinguishable from situation B.

Your empirical validation proceeds as follows: You note a location where the particle frequently comes to rest. (It is natural to assume that this location is a singularity--a stable location of minimal potential energy--in the particle-surround system.) You then arrange matters so that on the next occasion that the particle is immobile at that location, the entire surround moves relative to the particle. You observe that the particle displaces in the same direction as the surround.[10] You conclude that the vector flow field lawfully generated by the displacement of the surround in direction +X specifies a displacement of the particle from the singularity in direction -X. Hence, the particle displaces in direction +X toward the singularity.

This kind of kinematic closure differs from the more familiar types. The two familiar types you have already remarked upon might be labeled (1) kinematic closure through resistant bodies and (2) kinematic closure through forces. The kinematic closure you are now promoting is (3) kinematic closure through specification. The three types are alike in that the realization of any particular motion requires that a special relation hold between the paired elements. You are convinced that if you were observing your particle on a rectilinear trajectory toward a given region of the surround and you intruded on the flow field by some means so as to introduce a prolonged rotational component into the flow field, then the rectilinear trajectory would not be maintained. To realize any given trajectory of the particle, a symmetry must exist between that trajectory and the flow field: For the particle to move clockwise there must be a counterclockwise flow; for the particle to move toward p there must be a flow centered at p, and so on. Although it is very clear to you that for your particle and its ambient field this symmetry always holds, the point that you wish to underline is that in the absence of this symmetry an "intended" trajectory cannot be satisfied.

You are absorbed by what the foregoing reasoning implies, namely, that there might well be a similitude for all machines, hard-molded and soft-molded. The invariant feature of machines seems to be kinematic closure achieved by reciprocal contexts of constraint; kinematic closure seems to be founded on a symmetry between the paired elements. To your journeyman understanding, this symmetry reads: There is a transformation T such that if A and B are the paired elements, then T(A) ---> B and T(B) ---> A. You recognize that this transformation T is the mathematical notion of a duality operation and that the elements A and B are mathematical duals. You pose the question: What is the significance of the duality nature of machines? Tentatively you answer that if the prerequisite for constraining forces to produce selective, determinate motions is a duality structure, then duality must be a symmetry property of the most basic kind.[11]

Conclusion 12: In controlled collisions the particle must produce changes in force that are commensurate with changes in the kinematic field. Two examples come to mind: (1) to effect a soft collision any fluctuations in Pi(contact) that carry this quantity below its critical value must be countered by fluctuations in the control quantity, C, that are of commensurate amplitude; (2) if the surround is caused to fluctuate, so as to produce oscil-

latory global outflow and inflow of the kinematic field, the particle's position will similarly fluctuate, 180° out of phase.[12] The particle's commensurate fluctuations are the result of force changes in proportion to flow changes.

Your earlier conclusions about the conditions of the coupling of particle and field are incomplete. They do not identify a principled physical basis for force differences that are proportional to flow differences. When considering hydrodynamic flow, you normally visualize a process in which an inhomogeneity in potential gives rise to a force that drives a flow. More generally, differences in potential (ΔP) give rise to differences in force (ΔF) that, in turn, give rise to differences in flow (ΔV):

$$\Delta P \text{ -----> } \Delta F \text{ -----> } \Delta V.$$

Flows are proportional to forces, and where the Onsager condition holds, sensible deductions can be made in many instances from the macroscopic hydrodynamic flow to the irreversible thermodynamics that is its basis. The problem your particle poses is different from this conventional problem. It reverses the causal path and asks how flows can give rise to proportionate forces. Here, the causal vocabulary looks strained. But you are aware that you have felt this strain throughout your analysis. Thus you have spoken of the kinetic fields (particle and surround) as causing the kinematic field and the kinematic field as specifying and, cognately, constraining the kinetic field.

You remind yourself of some basics: Changes in motion or flow per se cannot cause changes of force; there can be no forces where there are no potential differences; the trajectory of force depends on the form of the potential. You surmise that if a flow is to affect a force it must do so by modifying the potential from which the force is derived. Modulating a potential would not necessarily cause a change of force; generally, other conditions must be satisfied. This reservation is consonant with your observation of the influence of the flow field on the particle: only global changes in flow lead invariably to changes in force. So, a change in force may or may not occur given a change in flow, but what you are after is a lawful basis for these changes whenever they do occur.

The problem has been refocused: How could a flow affect a potential? Formally, a force F is defined as the negative of the potential inhomogeneity or, more precisely, gradient, viz.,

$$F = -\nabla P,$$

where the gradient symbolized by ∇ is a spatial gradient. If P is identified as the particle's on-board potential, which is taken to be nearly homogeneous (given the arguments you made about the reversible, close-to-equilibrium steady state of the particle--Conclusion 9), then you must look to the kinematic field as the source of the inhomogeneity, that is, as specifying a spatial operator, ∇. Now, by taking the first derivative of both sides of the above expression for F you get

$$dF/dt = -d(\nabla P)/dt;$$

that is, control (see Conclusion 1) is given by the rate of change of the product of the spatial operator and the potential. In the foregoing context the first derivative of $-\nabla P$ defines a temporal gradient. As with the spatial gradient, you take the temporal gradient to be an operator defined by the kinematic field. Assuming commutativity the preceding expression for the control quantity can be written

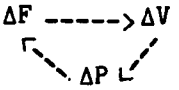
$$dF/dt = - \nabla dP/dt = -\partial^2 P/\partial X_i dt,$$

where ∂X_i is the spatial operator and dt is the temporal operator. In sum, the answer to the question of "how could a flow affect a potential?" seems to require the recognition and understanding of space and time operators on potentials. Given that the units of space and time must be in the scale of the particle--expressed in terms of the mean free path δ and the mean relaxation time τ of the particle's interior--the control quantity ought to be reducible to an expression in P, δ changes and τ changes.

As a further point, the ordering of potential, force, and flow that you are suggesting here is different from that which follows from considerations of hydrodynamic flow, namely:

$$\Delta V \text{ -----> } \Delta P \text{ -----> } \Delta F.$$

It would be prudent, however, to relate the two orderings. You go for the most obvious relation:



The flow field (ΔV) and energy flux (ΔP -----> ΔF) are linked in "circular causality." You underscore that these two "paths" of influence are not the same. First, the flux to flow path is a change in layout (e.g., a flow is produced when the particle as a unit displaces relative to the layout of the surrounding regions) whereas the flow to flux path is through the translational gate you identified in Conclusion 9. Second, comparatively speaking, the flux to flow path is energetically expensive, whereas the flow to flux path is energetically cheap (see Conclusion 9). (You resist identifying these paths with the cybernetic notions of "forward fed" causality and "backward fed" causality. You feel that such a move is regressive given that the notions of feedforward and feedback imply a referent signal, a comparator and, more generally, a separate controller. The origin and functioning of each of these would have to be rationalized by physical principles. [As a physicist you wish to explain the phenomenon of controlled collisions without the introduction of controllers sui generis.] Moreover, you feel that the different labeling of the pathways, as forward and backward, while well-motivated in artifactual situations, is arbitrary in natural situations.)

Conclusion 13: A controlled collision is a physical event in space-time. It is, however, by the conventional theory of physical events, a very odd kind of event. You struggle to formulate its heterodox quality: A controlled collision is a space/time event in which the final conditions of a particle's motions determine the values that the initial conditions must assume. (You

had observed repeatedly, for example, that when the particle softly collided and when it violently collided with a region of the surround, its accelerative change prior to collision was initiated at two different marginal values of the time-to-contact property.) This heterodox quality suggests to you a structure of space-time peculiar to controlled collisions, one that is explicitly shaped by both initial and final conditions. As a physicist you are well aware of the need to be clear on the space-time structure of events. Without a prescription for putting space-time boundaries on an event, the determination of its causal basis remains very much a guessing game. Within what limits should you try to close the bookkeeping on the relevant summational invariants--the conservations? You turn your attention to conventional physical event theory to see how well it fares in this regard and to see what modifications will be required.

In the conventional theory, "observer" refers to the measurement of the location of an event in space-time. As a local reference system or inertial frame, the observer must be perspective free. Measurements must be made simultaneously and distributively throughout a given region of space-time. The "observer," therefore, must be capable of existing everywhere in a specified region of space-time. Your particle "observes" and "measures" (its surroundings and its relation to them). However, given that it is of finite size (rather than being infinitely small) and can exist in only one place at any one time, it cannot be identified with the observer in orthodox physical event theory: Your particle must have a perspective. You suspect that this fact will be of importance in the eventual formulation of the laws of controlled collisions.[13]

Corollary to the absence of a real or natural perspective in physical event theory is the absence of an historical perspective. While the present is causally constrained by the immediate past, it is not (to borrow a term from Bertrand Russell) mnemically conditioned by the distant past. You sketch for yourself the Minkowskii diagram (Figure 7) that illustrates the causal light cone that is the traditional domain of physical event theory. (Figure 7b is a simplified version of Figure 7a, with x, y, z reduced to a single spatial (s) axis.) With the speed of light as the limiting boundary, only those events within the same forward light cone can be causally connected to the present event at the origin, t=0 (because there are no known superluminal signals, events outside the light cone cannot be connected with those inside). The events leading up to the present are nowhere represented. The premise of the orthodox theory is that the past is instantiated in the present and that, together with the laws of motion, is sufficient to predict or explain event outcomes. The particle you have been studying makes you skeptical of this premise. Somehow the final conditions must be brought in--explicitly--to accommodate controlled collisions.

You try to close in on what this would require by producing a series of modifications of the Minkowskii diagram. First, you include a past light cone that converges at t = 0--the event from which the forward or future light cone diverges. In your modified sketch (Figure 8) you have rotated the axes so that time flows from left to right. Next, you depict four events in your sketch (Figure 9). The events E_1, E_2, and E_3 are on the same world line where E_3 is causally constrained by E_2 and E_2 is causally constrained by E_1. You take pains to note that the causal constraints are not necessarily transitive for these interactional sequences (that is, E_3 is not necessarily causally constrained by E_1). This is because E_2', which is on a world line with E_3,

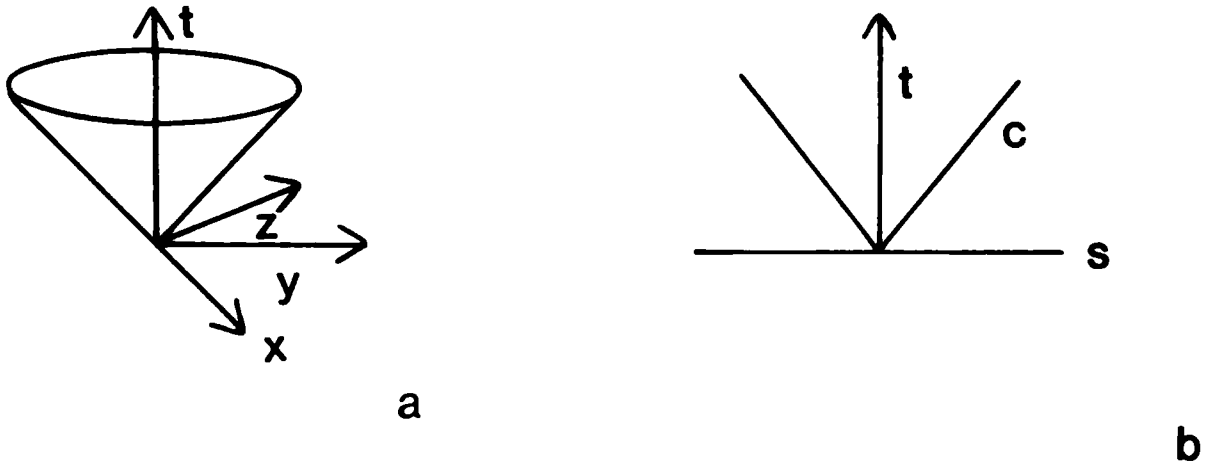


Figure 7. (a) The causal light cone determined by time (t) and three spatial dimensions, x, y, and z. (b) The causal light cone where x, y, and z have been reduced to a single spatial axis (s), showing the speed of light, c, as the limiting boundary.

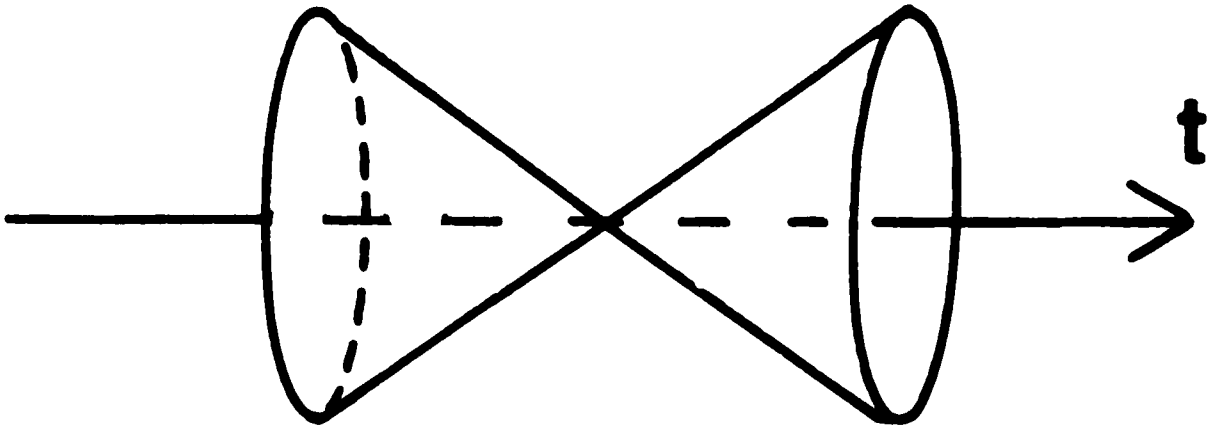


Figure 8. A modified Minkowskii diagram rotated so that time flows from left to right. It includes a mnemic (past) light cone as well as the standard causal (future) light cone.

might cancel (or otherwise alter) the effects of E_1. While E_1 transacts with E_3 in the context of E_3's historical relation to E_2, it does not do so in terms of the historical context of E_2. The rub, as you see it, is that because E_2' is outside the forward light cone of E_2 (it is effectively simultaneous with E_2), its effects cannot be known at E_2 and, therefore, E_3 cannot be explained on the basis of E_2's causal cone alone.

Because unobservable events may exert an influence on future events, necessary paths of influence cannot be discovered by working forward from initial conditions to final conditions. You recognize, however, that determinant histories may be discovered by working back from the final conditions to the initial conditions. All of the influences on E_3 are in its past or mnemic cone. In sum, the causal future of E_2 is only partially accounted for by its forward cone but all of the determiners of E_3 are in its mnemic cone. There is an asymmetry between the information derived from history and the information applicable to the future.

You are inclined to believe that the only appropriate framework for controlled collisions must be composed of the causal and mnemic perspectives together. But is this framework to be one in which these perspectives remain asymmetric? Or, more accurately, is there a different level of analysis that may reveal the symmetry of the event space for controlled collisions? You pose this question because of a major lesson learned from orthodox physical event theory: Putting symmetry at the forefront reveals the structure of space-time and fetters the application of law. Knowing the symmetry that defines a space-time event means that if one element of an event is known, the nature of its symmetric counterpart is also known.

You modify your sketch of the Minkowskii diagram once again, this time creating a bounded region between the causal and mnemic cones of two succeeding events (Figure 10). You are now ready to propose a symmetry postulate for controlled collisions: If (i) E_1 (approach to a region) and E_2 (contact) are on the same world line (where E_2 is in the causal cone of E_1 and E_1 is in the mnemic cone of E_2) and (ii) there are no events outside the causal cone of E_1 that influence E_2, then E_1 and E_2 together define a new event--call it E_D--for which they are dual perspectives. The past and future cones have been merged into a higher order event space. Events outside the bounded region have no existence for the particle; they are in neither its history nor its future. Events inside the bounded region have relative existence. The new event E_D is a controlled collision and it will be guaranteed whenever the symmetry conditions (i and ii above) hold.

In a further sketch (Figure 11) you contrast dual events with non-dual events. The events E_0 and E_1 are duals, the events E_2 and E_3 are duals, but E_1 and E_2 are not duals because condition (ii) is violated (E_2 is influenced by E_1' which is in the null cone of E_1). What you wish to show in this last sketch is that the specification of E_2 will be indeterminate when based on the causal cone perspective of E_1. Moreover, the selection of marginal values at E_1 to determine an outcome at E_2 is not guaranteed to be successful since the basis for controlling the outcome at E_2 is not completely available at E_1. A controlled collision cannot be defined over E_1 and E_2 because they are not duals.

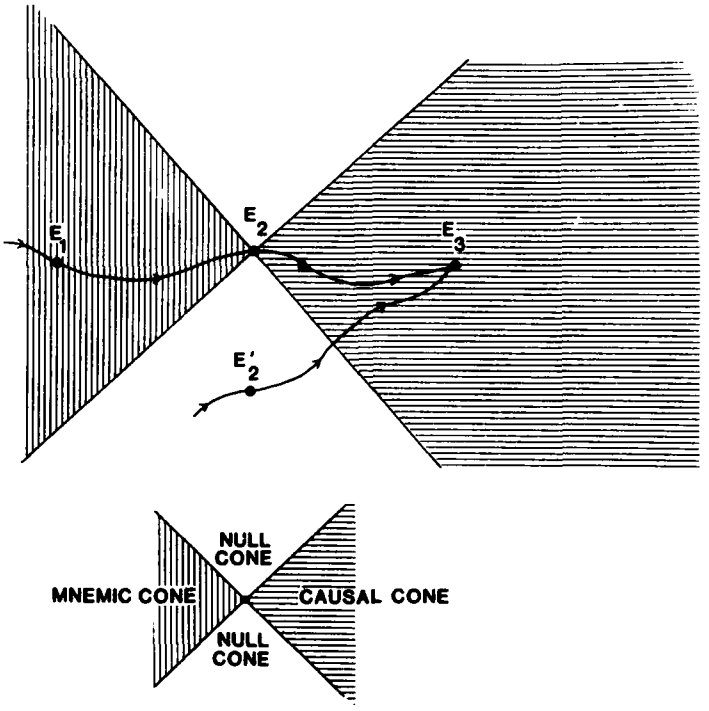


Figure 9. The causal relationships among four events. Although E_3 is in the causal cone of E_2, it cannot be explained on this basis alone--E_2' exerts an influence on E_3, yet is in the null cone of (and, therefore, unknown at) E_2.

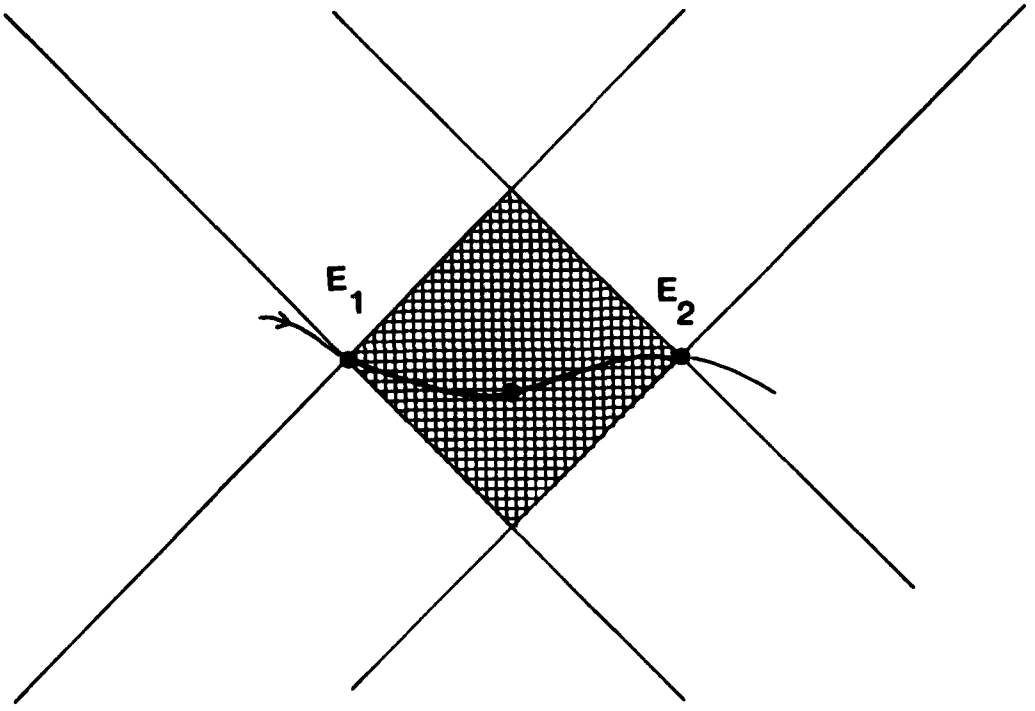


Figure 10. The bounded region between the causal cone of E_1 and the mnemic cone of E_2 defines a new event, E_D, for which E_1 and E_2 are dual perspectives.

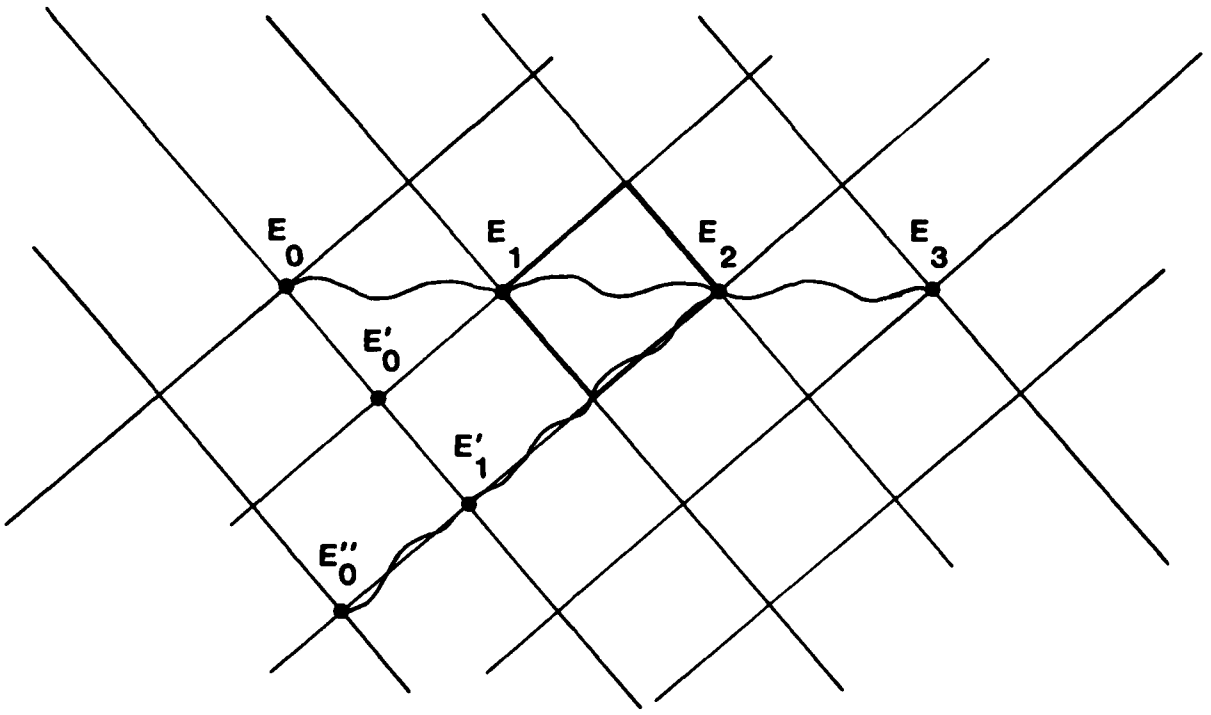


Figure 11. E_0 and E_1 are duals (note that E_0'', though in the null cone of E_0, is not on a world line with E_1), as are E_2 and E_3 (note that E_0'' is at the limiting boundary of (and, therefore, is included in) the mnemic cone of E_2). E_1 and E_2 are not duals because E_1 influences E_2 but is in the null cone of E_2.

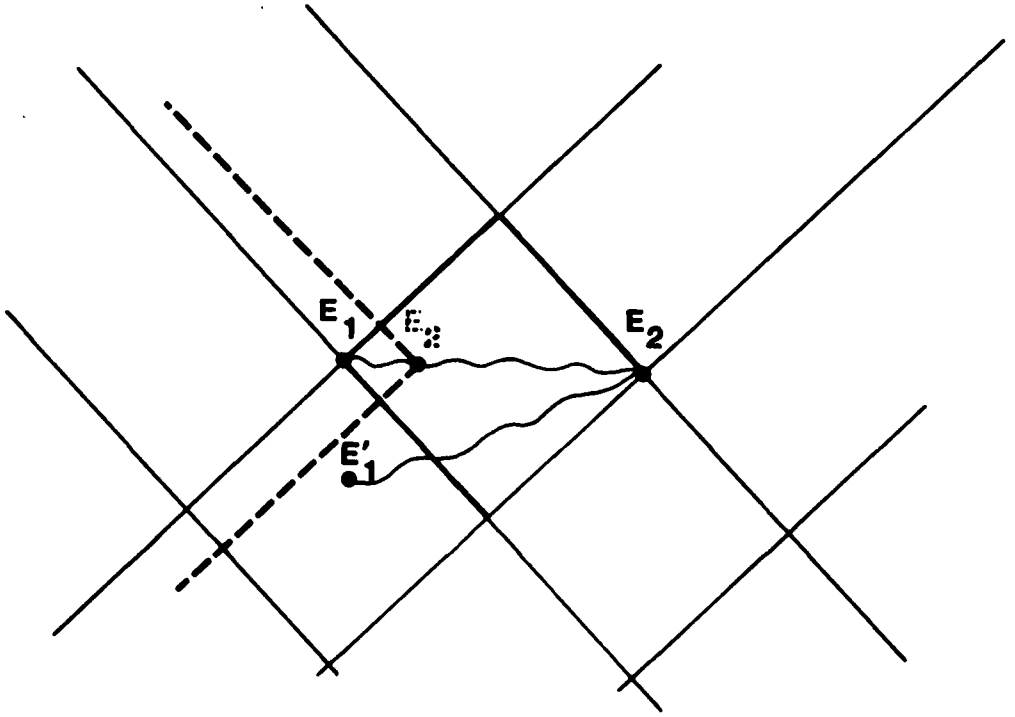


Figure 12. Some E_2 must exist that is causally proximal to E_1. A change in scale reveals the duality over which a controlled collision can be defined.

To restore or, more accurately, to reveal a duality, you suggest a change in scale (Figure 12). At the grain of a finer space-time mesh there necessarily exists some event E_2, causally proximal and dual to E_1, for which a controlled collision can be minimally defined. This change in scale merely assumes that the particle has limited sensitivity or acuity to distant events on its world line. (In fact, your observations of many particles of varying sizes reveal that there is a strong relationship between acuity and size. The spatial range is a constant proportionality of the vertical magnitude of the particle.[14] Simply put, large particles act with respect to things at a greater absolute distance than do small particles.)

Your point is that for controlled collisions, any events antecedent to some future event toward which the particle's current behavior is directed (1) must lie within the particle's current causal perspective if they have significant effects on the particle's immediate future or (2) must be trivial in their effect if they lie undetected in the particle's null cone. Because significant events cannot lie outside the bounded region of a controlled collision, an appropriate scale of analysis that satisfies this condition must exist. You insist that symmetry is the guide to finding this scale: Given either the perspective from the initial conditions or the final conditions, the other perspective is specified.

A Summary and an Awakening

You have discovered quite a lot about your particle, but its identity still eludes you. You convince yourself that you have all the information you need to identify this type of particle and it is only some firmly entrenched bias that prevents you from seeing it. You think that you may have given a physical description to the behavior of an entity that is usually considered to be outside the domain of physics. Several of its properties are like those of more standard particles, but you have noticed they often include less standard twists. You review the properties you have discovered in the hope that highlighting the "twists" might fuel an insight. (At the very least, it will provide a convenient way to summarize these REM episodes.)

(1) The behavior of your particle can be described with a measurable quantity but this quantity, is control ($\Delta MV/T^2$) rather than the more standard momentum (MV).

(2) Forces determine the trajectory of your particle, but they are dissipative rather than conservative forces and they originate not in the surround but in the particle. Moreover, the particle can replenish its energy supply.

(3) The distribution function that you constructed as a means of classifying your particle reveals it to be in a class whose behavior is not governed by velocity-dependent conservations.

(4) Your particle exhibits conservation, but it seems to be conservation of population number, rather than the more standard energy or momentum or mass. To accomplish this conservation, it appears to minimize momentum transfers that might fracture the particle.

(5) Because your particle harnesses forces to achieve selective trajectories, you consider it to be in the class of machines. But its constraints are soft-molded, allowing flexibility in the strength of collisions, rather than hard-molded.

(6) The soft constraint on the particle-based forces is a field, but it cannot be associated with a force.

(7) Because the constraining field is not a force field, it cannot include dimension M and, therefore, is not kinetic; because certain properties that are necessary to the control of collisions are annihilated when t goes to 0, the field must include dimension T and, therefore, is not geometric. The soft-constraint must be a kinematic field.

(8) Critical values of dimensionless quantities in the kinematic field distinguish between qualitatively distinct states, but these are specificational states rather than physical states as would be the case in a kinetic field.

(9) Because the kinematic field constrains the particle's trajectory, it must be coupled somehow to the particle, but the coupling must be linear (so that equivocalities are not introduced) and low energy (so that it does not involve large momentum exchanges and irreversible processes).

(10) You explain the coupling through a conservation, but it is of topological form (adjacencies and successivities) rather than of energy or momentum.

(11) The machine conception (identified in Conclusion [4]) must apply minimally to the particle and the field as duals, not just the particle. The symmetry is necessary in order to realize and maintain trajectories.

(12) The flow field produces proportionate forces in the particle, presumably by modulating a layout of potentials. Whereas the fact that forces produce flows proportionate to the forces is understood, the fact that flows produce forces proportionate to the flows is not.

(13) Controlled collisions, which are characteristic of your particle, are physical events, but the structure of space-time is shaped by final conditions as well as initial conditions. Where the particle is going colors how it gets there.

What is this soft-coupled duality of particle and surround, wherein collisions are guided by distinct specificational states that bring final conditions to bear on initial conditions, and are controlled by the dissipation of the particle's replenishable energy reserves in such a way as to minimize momentum transfers that could fracture it? You seem to have described a physics of controlled collisions, but for what...or whom...?

You are startled awake by the agitated chirping outside your window. The bird is hovering about a feeder in an effort to replenish its fuel supply, but a cat has appeared on the scene waiting to replenish itself by effecting a violent, predatory collision on your friend. Fortunately for the bird, you muse pretentiously, the imminence of contact with the cat is specified in the optical flow field that links properties of the animal to properties of the

environment. You marvel, once again, as it guides its flight to avoid the cat and locate the food, cutting its speed just in time to alight gently on the feeder. Now those are the kinds of controlled encounters that Gibson wanted to understand and that you've been trying to understand. You are suddenly overcome with a sense of déja vu, with a feeling that, at some level, you have understood.

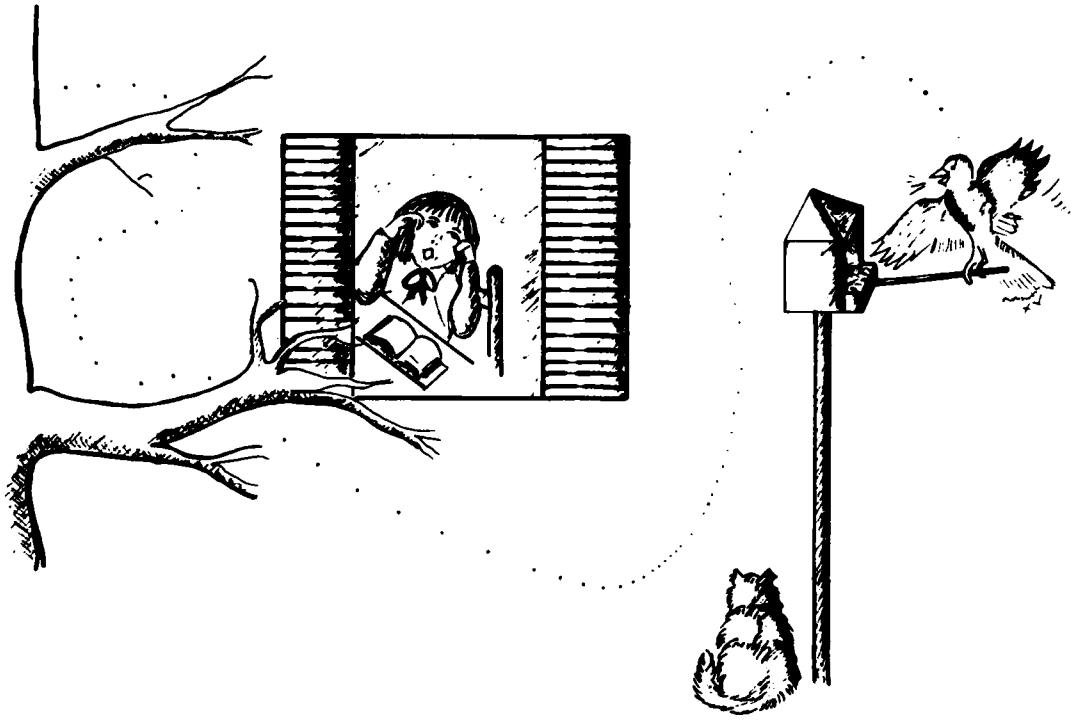


Figure 13. The dreamer awakes.

References

Buckingham, E. (1914). On physically similar systems; illustrations of the use of dimensional equations. Physical Review, 4, 345-376.

Freudenthal, A. M. (1950). The inelastic behavior of engineering materials and structures. New York: Wiley.

Gibson, J. J. (1950). The perception of the visual world. Boston: Houghton-Mifflin.

Gibson, J. J. (1960). The information contained in light. Acta Psychologia, 17, 23-30.

Gibson, J. J. (1961). Ecological optics. Vision Research, 1, 253-262.

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton-Mifflin.

Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton-Mifflin.

Goldsmith, W. (1960). Impact: The theory and physical behavior of colliding solids. London: Edward Arnold Ltd.

Hoyle, F. (1957). The black cloud. London: Heinemann.

Iberall, A. S. (1977). A field and circuit thermodynamics for integrative physiology: I. Introduction to general notions. American Journal of Physiology/Regulatory, Integrative, & Comparative Physiology, 2, R171-R180.

Kirschfield, K. (1976). The resolution of lens and compound eyes. In F. Zettler & R. Weller (Eds.), Neural principles in vision. Berlin: Springer-Verlag.

Kornhauser, M. (1964). Structural effects of impact. Baltimore, MD: Spartan Books Inc.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. Perception, 5, 437-459.

Lee, D. N. (1978). The functions of vision. In H. Pick, Jr. & E. Saltzman (Eds.), Modes of perceiving and processing information. Hillsdale, NJ: Erlbaum.

Lee, D. N. (1980). Visuo-motor coordination in space-time. In G. E. Stelmach & J. Requin (Eds.), Tutorials in motor behavior. New York: North-Holland.

Lishman, J. R., & Lee, D. N. (1973). The autonomy of visual kinaesthetics. Perception, 2, 287-294.

Nadai, A. (1950). Theory of flow and fracture of solids, Vol. II. New York: McGraw-Hill.

Pattee, H. H. (1972). Laws and constraints, symbols and language. In C. H. Waddington (Ed.), Towards a theoretical biology. Chicago: Aldine.

Pattee, H. H. (1977). Dynamic and linguistic modes of complex systems. International Journal of General Systems, 3, 259-266.

Reuleaux, F. (1963). The kinematics of machinery. New York: Dover.

Schuring, D. J. (1977). Scale models in engineering. New York: Pergamon Press.

Shaw, R., & Turvey, M. T. (1981). Coalitions as models for ecosystems: A realist perspective on perceptual organization. In M. Kubovy & J. Pomerantz (Eds.), Perceptual organization. Hillsdale, NJ: Erlbaum.

Timoshenko, S., & Goodier, J. N. (1951). Theory of elasticity. New York: McGraw-Hill.

Walton, A. J. (1976). Three phases of matter. New York: McGraw-Hill.

Warren, W. H., Jr. (in press). Perceiving affordances: The visual guidance of stair climbing. Journal of Experimental Psychology: Human Perception and Performance.

Yates, F. E. (1982). Outline of a physical theory of physiological systems. Canadian Journal of Physiology and Pharmacology, 1982, 60, 217-248.

Footnotes

[1]Iberall (1977) has suggested that the number of members of a biological species is approximately conserved and a physics that accommodates biology will require the addition of this conservation to the list of conventional conservations.

[2]Gibson's optic array (1961, 1966, 1979) seems to be a field of this type.

[3]Gibson repeatedly pointed out that optical motion is altogether different from material motion--that optical motion has no inertia (for example, Gibson, 1979).

[4]Properties of this kind were identified by Gibson (1966, 1979) for the optical flow field resulting from the locomotion of an animal in a cluttered environment.

[5]Lee (1976, 1980) identified this property for the condition in which a point of observation approaches, or is approached by, a substantial environmental surface.

[6]See Gibson's (1979) discussion of the optical support for the control of locomotion.

[7]Lee (1976, 1980) performed these calculations and highlighted the significance of the first derivative of the time-to-contact variable. Other optically defined dimensionless quantities that order (at critical values) specificational states have been suggested and experimentally examined by Warren (in press).

[8]For animals, the photoreceptor processes perform the role of a translational gate that involves very little energy relative to the animal's daily energy expenditure.

[9]Such as Reuleaux (1963).

[10]Lishman and Lee (1973) have shown that in a room where the walls and ceiling can move as a unit, displacement of the room causes a person standing in the room to topple in the direction of the room's movement.

[11]This point has been argued by Shaw and Turvey (1981).

[12]See Lee (1978).

[13]For Gibson (1966, 1979) the structure of an optical flow field is always exterospecific and propriospecific--it is always specific to the layout and to the observer.

[14]Kirschfield (1976) reports that for animals there is a simple first-order relation between visual resolution (R) and body-height (H), $R = k/H$, where k is a constant of proportionality.

Appendix

A. The Theory of Collisions

The concept of collision refers to forces applied to and removed from an object in a very short period of time. The classical theory of collision, based primarily on the impulse-momentum law for rigid bodies, regards the colliding objects as single mass points. All elements of each object are assumed to be rigidly connected and to be subjected instantaneously to one and the same change of motion as the result of the collision. In reality, the forces initiate stress waves that travel at finite velocity away from the region of contact and through the object. These waves reflect from boundaries of the object and interact with stress waves still being generated at the region of contact to create a complex pattern of stresses and strains in the interior. In short, all regions of an object subjected to a collision are not exposed simultaneously to the same force conditions (Goldsmith, 1960).

The classical theory is most suited to ideal atomisms whose degrees of freedom are exhausted by the three axes of translation. Atomism is a term suggested by Iberall (1977) for an entity of any magnitude that is atom-like at the scale of the ensemble to which it belongs. It is conventional to say that ideal atomisms have no internal degrees of freedom, where "internal" has the uncommon meaning of "extra-translational." Atomisms of gases such as helium are closest to this ideal. They are single atoms each free to move on the three spatial dimensions. For all intents and purposes, the total energy imparted by collision to a helium atomism may be regarded as going into the translation of the atomism. In terms of the equipartition theorem, the energy received is divided evenly and completely among the atomism's degrees of freedom, which are all translational.

The atomisms of another gas, oxygen, introduce a measure of internal complexity. These atomisms (molecules) consist of two linked atoms. To define the position of each of the atoms of oxygen requires three degrees of freedom, for a total of six. However, the linkage between the atoms eliminates a coordinate choice, thereby reducing the degrees of freedom of the oxygen atomism to five. Because translation of the oxygen atomism's center of mass consumes only three of the five degrees of freedom, the two degrees of freedom that remain are "internal." The equipartition theorem would assign three-fifths of the energy of collision to the translation of the atomism and two-fifths of the energy to the internal bond. Clearly, conservation of energy does not hold if only the energy carried by the translational degrees of freedom is taken into account. It is for this reason that collisions of atomisms with internal degrees of freedom are said to be inelastic and that the conservation of momentum (rather than of energy) is the dominant constraint on their equations of collision.

Consideration of the collisions of di-atomic atomisms is a small step toward the collisions of systems. In a statistical mechanical sense a system is an ensemble of interacting atomisms with a boundary that prohibits the ensemble from dissolving into the surround. The atomisms of a system may be internally barren (like the helium atomism) or internally complex (of a kind hinted at by the oxygen atomism). As noted, internal complexity is associated with ways of absorbing the energy applied to a unitary thing other than through the translation of its center of mass.

B. The Theory of Fracture

The first major advance beyond the classical theory of collisions (viz., the one-dimensional vibrational treatment of colliding objects) recognized the significant proportion of energy converted into oscillations when the system's natural frequency is long compared to the duration of contact. Subsequent analyses of the multi-dimensional aspect of wave propagation consequent to collision, and of the stress distribution at the region of contact, were made possible by developments in the theory of elasticity (Timoshenko & Goodier, 1951). It suffices to say, for present purposes, that elasticity refers to the fact that the internally generated forces of restoration are comparable to the externally applied forces of deformation so that there is a return to the status quo ante on removal of the external forces.

In many collisions, however, the conditions of impact are such that the entire cross-section of one or both of the colliding objects will exhibit a final permanent strain of significant magnitude, or one or both of the objects

may fracture. Such non-reversible phenomena result from the conversion of kinetic energy into permanent distortion or fracturing of the structure of the object and the eventual dissipation of this energy in the form of heat. The analysis of the irreversible deformations wrought by the propagation of stresses that exceed the elastic limit (so called plastic flows or plastic waves) is a more recent and less developed aspect of collision theory (Goldsmith, 1960).

Evidently, the responses of an internally complex system to collision will be difficult to follow. It is possible, nevertheless, to obtain some useful insights into the collision process by considering (a) the behavior of a system under statically imposed forces and (b) the relation between impact parameters and system failure, ignoring the internal responses.

The deformation resulting from loading a system statically can be treated as a series of equilibrium states requiring no consideration of acceleration effects or wave propagations. Of major interest is the response to static loading of systems that exhibit a degree of rigidity, that is, systems that preserve their form in the face of perturbations. The requirement, of course, is that the system be elastic through some range of perturbation. Solids have an elastic domain as do multiphase systems that are solid or gel in part, such as living things that are dominated by elastic-plastic-fluid (liquid and gel) processes (Yates, 1982).

The interior of a solid system can respond in one of three ways to an applied force: (1) the linked atomisms can be forced further apart or closer together than the equilibrium (minimal potential) distance; (2) atomisms can hop into adjacent vacant lattice sites; and (3) the bonds between the atomisms can be broken (Freudenthal, 1950; Nadai, 1950; Walton, 1976). If (1) is sufficient to absorb the energy of loading, the solid is operating strictly within its elastic domain. Suppose that a static loading is realized as a force applied along an axis (a stress) so as to stretch or compress (more generally, to strain) the system. Then response (1) means that the system as a whole undergoes a coordinate transformation that changes the distances between all the atomisms but not the topology of the system's internal configuration. This response to static loading is reversible. It is, however, a response of finite capacity. At some point the potential energy stored up within the excessively strained bonds reaches a limit (the elastic yield) and new mechanisms for accommodating the applied energy must be found (that is, a new "escapement" arises). One escapement mechanism is the breaking of some bonds between some atomisms (response 3), another escapement is diffusion (response 2) which is enhanced considerably by the structural changes resulting from bond breaking. (In a multiphase system at the elastic limit there is a structural change in at least one phase; for example, in the continuous solid phase of a two phase solid-fluid system such as a gel or in the more rigid phase of a polyphase solid-solid system such as a polycrystalline metal or a polyphase solid-fluid phase system such as a high polymer.)

A brittle system (a physical ideal, an engineering myth) would fracture at the elastic limit. There are no plastic deformations (flow processes) in a brittle system and microscopic bond breaking becomes, immediately, macroscopic fracture. For real, ductile systems, however, the yield point only identifies that loading at which fracturing begins on the atomistic level. Once the yield point is reached in a ductile system, the mutually reinforcing processes of bond breaking and diffusion can continue to accommodate excessive energy

brought in by the static loading. The dissociating of some of the atomisms makes it easier for other bound atomisms to migrate to locations that are more stable than the locations that they currently occupy. This flow process is irreversible: Less energy is required for an atomism to hop from a high to a low potential site than vice versa. However, the consequent relaxing of some bonds brought about by diffusion increases the strain on other, already overstrained, bonds, disposing them to further fracture.

The micro-fracturing that begins at and proceeds beyond the yield point reduces the long range order or cooperativity of the system (interpreted as bonds that repeat regularly over many thousands of atomic distances). The long range order is replaced by short range order or local cooperatives, not unlike the "flow unit" of a liquid. The diffusion occurs at the surface of these local clusters because the atomisms located there are thermally less stable than their partners in the interior. Clearly, the larger the number of local cooperatives and, therefore, internal surfaces, the greater the diffusion. And the greater the diffusion the more disposed to breaking are the already strained bonds at places in the system where diffusion of atomisms is not possible. In sum, fracturing of the bonds between atomisms is a chain reaction process and eventually a ductile system will fracture at the macroscopic scale.

The emphasis of the foregoing has been the gradual progression of macroscopic fracture, or system failure, as might occur under the repeated or prolonged application of static forces that exceed the system's elastic limit. In the range between the initiation of bond breaking on the microscale and the occurrence of system failure on the macroscale, the system gradually loses its ability to absorb the applied energy. A measure of the energy absorption of a material is given by its stress (force per unit area)-strain (proportional change in length) curve. The energy per unit volume is approximately equal to the shaded area of Figure 14. Consequently, the strain energy to failure may be approximated as follows: energy/unit volume = $1/2\ (\rho_x + \varepsilon_x)\rho_c$. Where ρ_c is the stress at the yield point and ρ_x and ε_x are the ultimate stress and ultimate strain, respectively, that mark the collapse of the system.

Of course, the loss of the ability to absorb energy could be quick, given a collision. The microscopic processes leading to failure from a single brief loading must be a rapid chain reaction of bond breaking associated with elastic and plastic waves propagating from the point of contact and multiply reflecting from the system's boundary. However, as noted, broad conclusions relating failure to the conditions of collision are possible without considering the complex of intermediary processes.

A collision will have an acceleration (of the system) X time profile. Three examples of single loadings are given in Figure 15; to achieve a given response amplitude, shorter durations of loading must be compensated by greater accelerations. Two parameters are of special significance: the change in velocity and the average acceleration (in units of gravity) that is just sufficient to produce structural failure. In Figure 15 the cross-hatched areas express the velocity changes. The average acceleration of any collision is equal to the velocity change divided by duration. A collision sensitivity curve can be generated by plotting criterial velocity change (where fracture occurs) against criterial average acceleration (where fracture occurs) (Kornhauser, 1964). A prototypical collision sensitivity plot for a prototyp-

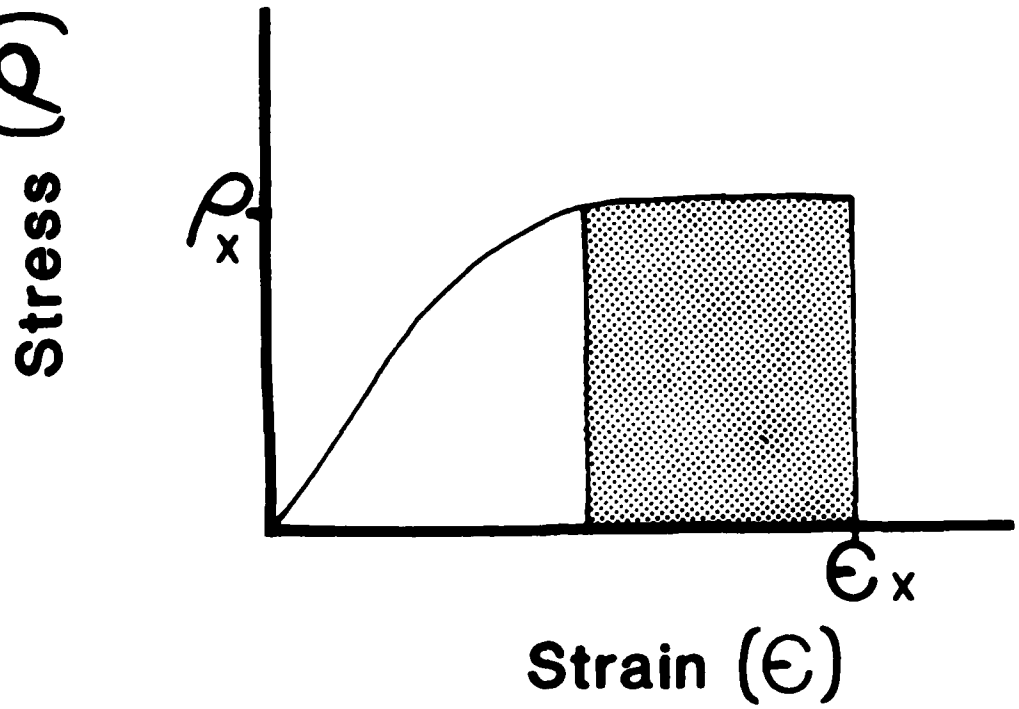


Figure 14. The energy absorption per unit volume of a material is given by the shaded area of its stress-strain curve.

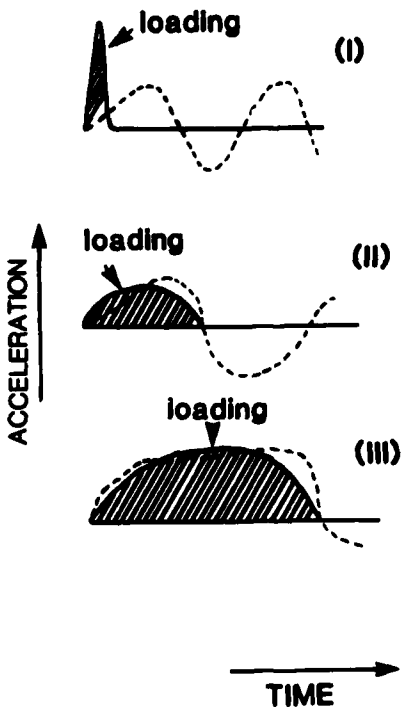


Figure 15. Acceleration x time profiles of collisions under three loading durations (after Kornhauser, 1964).

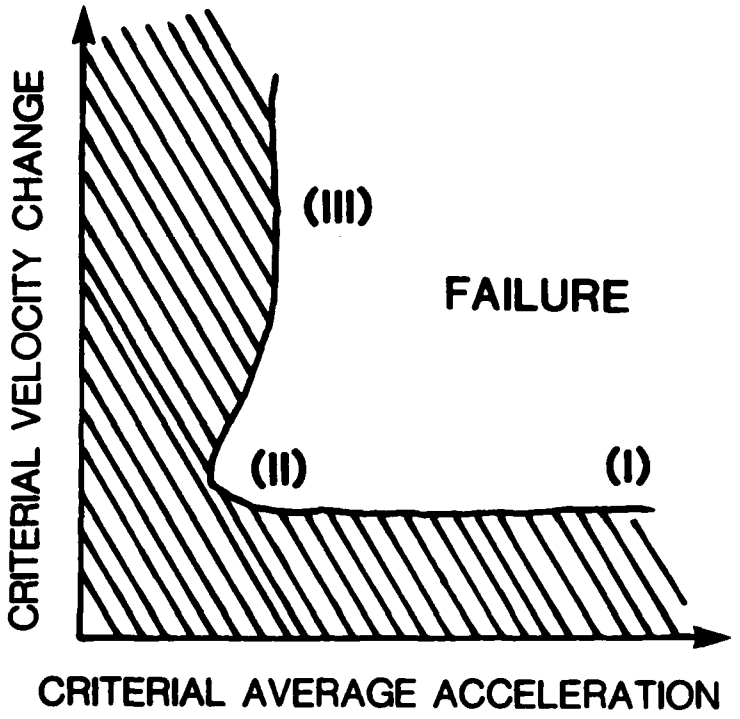


Figure 16. Collision sensitivity plot shows where system failure will occur (after Kornhauser, 1964).

ical system is given in Figure 16. The vertical asymptote is related to acceleration pulses that are steady or of long duration. It implies that no failure occurs unless a certain average acceleration is exceeded, regardless of the change in velocity of the system and the duration of the collision. The horizontal asymptote is related to acceleration pulses of short duration. It implies that system failure does not occur unless a certain velocity change is exceeded regardless of the average acceleration value (Kornhauser, 1964).

The location of the vertical asymptote in Figure 16 is a function of the shape of the collision (its acceleration X time profile). In contrast, the horizontal asymptote is independent of the shape of the loading and is fully characterized by a unique value of velocity change: Collision durations that are short enough to be on the short duration asymptote (marked by (I) and (II) for a given system will result in the structural failure of that system. There is some evidence (Kornhauser, 1964) to suggest that the collision velocity change required for irreversible damage to mammals is relatively indifferent to species and size (25 feet per second is a reasonable approximation). The criterial average acceleration, however, differs markedly with species and size (roughly, 20 g for man and 650 g for mice).

A simple rule of thumb relates the criterial velocity change (V_c) and criterial average acceleration (G_c) to the system's natural frequency (ω) (Kornhauser, 1964):

$$G_c = \omega V_c.$$

If most collisions between systems and their surrounds are of sufficiently short duration to place the systems on the horizontal asymptote of their collision sensitivity function, then V_c is constant. (For mammals, as noted above, V_c = 25 f/s.) In other words, the higher the value of a system's natural frequency, the greater is the system's tolerance to collision (measured in multiples of the gravitational constant).

ON THE PERCEPTION OF INTONATION FROM SINUSOIDAL SENTENCES*

Robert E. Remez† and Philip E. Rubin

Abstract. Listeners can perceive the phonetic value of sinusoidal imitations of speech. These tonal replicas are made by setting time-varying sinusoids equal in frequency and amplitude to the computed peaks of the first three formants of natural utterances. Like formant frequencies, the three sinusoids composing the tonal signal are not necessarily related harmonically, and therefore are unlikely to possess a common fundamental frequency. Moreover, none of the tones falls within the frequency range typical of the fundamental frequency of phonation of the natural utterances upon which sinusoidal signals are based. Naive subjects nevertheless report that intelligible tonal replicas of sentences exhibit unusual "vocal" pitch variation, or intonation. Our present study attempted to determine the acoustic basis for this apparent intonation of sinusoidal signals by employing several tests of perceived similarity. Listeners judged the tone corresponding to the first formant to be more like the intonation pattern of a sinusoidal sentence than either: (A) a tone corresponding to the second or to the third formant; (B) a tone presenting the computed missing fundamental of the three tones: or, (C) a tone following a plausible fundamental frequency contour generated from the amplitude envelope of the signal. Additionally, the tone reproducing the first formant pattern was responsible for apparent intonation even when it occurred in conjunction with a lower tone representing the fundamental frequency pattern of the natural utterance on which the replica was modeled. The effects were not contigent on relative tone amplitude within the sentence replica. The case of sinusoidal sentence "pitch" resembles the phenomenon of dominance, that is, the general salience of waveform periodicity in the region of 400-1000 Hz for perception of the pitch of complex signals.

Introduction

A number of recent studies of speech perception have examined the effects of sinusoidal replication of speech signals (Bailey, Summerfield, & Dorman, 1977; Best, Morrongiello, & Robson, 1981; Grunke & Pisoni, 1982; Schwab, 1981). Typically, such tonal analogs of speech are composed of three

---

*Also Perception & Psychophysics, 1984, 35, 429-440.

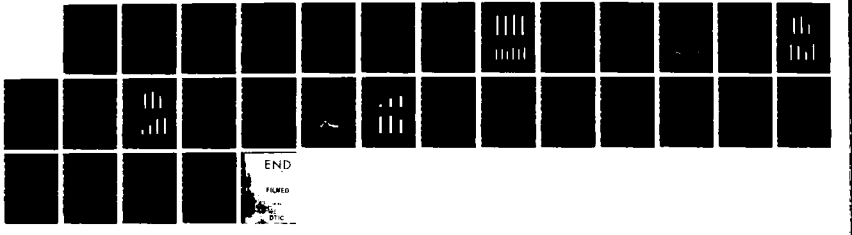
†Barnard College of Columbia University.

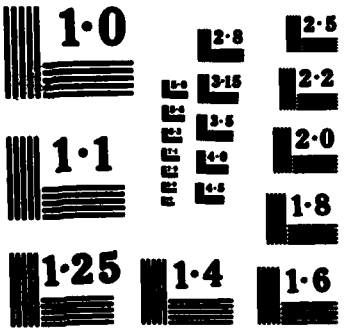
Acknowledgment. The authors thank Peter Balsam, Louis Goldstein, and Leigh Lisker for advice and encouragement. We are also indebted to our reviewers for insisting that we include Experiment 4 in the report. This work was supported by grants HD-15672 (to RER) and HD-01994 (to Haskins Laboratories).

[HASKINS LABORATORIES: Status Report on Speech Research SR-77/78 (1984)]

AD-A145 585 STATUS REPORT ON SPEECH RESEARCH A REPORT ON THE STATUS 3/3
AND PROGRESS OF S. (U) HASKINS LABS INC NEW HAVEN CT
A M LIBERMAN AUG 84 SR-77/78(1984) N00014-83-K-0083

UNCLASSIFIED F/G 17/2 NL




1·0
1·1
1·25
1·4
1·6
1·8
2·0
2·2
2·5
2·8
3·15
3·5
4·0
4·5

time-varying sinusoids, each tone reproducing the frequency and amplitude variation, sometimes schematically, of a formant from a natural utterance. In such acoustic patterns, devoid of the harmonic series and broadband formant structure of natural speech, the short-time acoustic properties are unmistakably not speechlike. Both acoustically and perceptually, sinusoidal signals are grossly unnatural, and naive listeners tend therefore to perceive sinusoidal sentences merely as several covarying tones unless they expect to hear a linguistic message; moreover, phonetic perception fails to occur unless the tonal stimulus is adequately structured, indicating that an explanation of this effect should be sought in terms of the information provided by these atypical stimuli (Remez, Rubin, Pisoni, & Carrell, 1981). When sinusoidal patterns are perceived phonetically, they are judged to be intelligible yet unspeechlike, presumably because they convey segmental information in an abstract pattern of spectrum variation, with almost none of the typical acoustic details of natural speech.

One consequence of this finding is methodological. This technique for transforming the signal can be used to reveal the perceptual significance of time-variation in the speech stream. This is so precisely because such unspeechlike signals disentangle the pattern of frequency variation over time in the speech stream from the sequence of particular momentary acoustic elements that are produced by vocal articulation. In view of the acoustic differences betweeen sinusoidal signals and the natural utterances that they replicate, it seems fair to suppose that sinusoidal replication does not merely reduce the amount of information present in the signal, as minimal-cue speech synthesis does (for example, Delattre, Liberman, & Cooper, 1955; and Abramson & Lisker, 1965). In that technique, a subset of the acoustic ingredients of an utterance is selected for imitating synthetically. Obviously, the information provided by natural acoustic elements is lost if those elements fail to appear in the synthetic replica. In such circumstances, phonetic information may or may not be adequately conveyed by the remaining acoustic structure. Therefore, this minimalist method is designed to reveal the effectiveness of particular acoustic elements--for example, a burst of noise, a low frequency murmur, or a prescribed frequency transition in the second formant--when others have been neutralized or eliminated.

In contrast, the transformation of a speech signal into time-varying sinusoids does not preserve particular constituents of the acoustic signal while discarding others. Rather, it destroys the physical similarity between acoustic moments in natural speech and those in sinusoidal patterns. The residual similarity between speech and sinusoidal imitations is to be found only in the variation of the two kinds of signal, and specifically in the pattern of frequency variation over time. For this reason, a significant aspect of the sinusoidal replication technique would be obscured by classifying the signals simply as "impoverished stimuli." They are, in fact, literal imitations of the time-varying properties of the supralaryngeal vocal-tract resonances. Sinusoidal signals of this type present the pattern of resonance center-frequency variation through an utterance, although the signals obviously do not contain formant structure.[1] Our tests (Remez et al., 1981) have established the sufficiency of this acoustic abstraction of the speech signal, in contrast to research that more customarily demonstrates the perceptual uses of selected brief pieces of the signal. When perceivers detect phonetic structure in sinusoidal patterns, this reveals the usefulness of the forms of stimulus change as phonetic information, and the independence of perception from most of the specific acoustic details with which the forms are conveyed.

Sinusoidal Intonation

In an obvious way, however, sinusoidal replicas of speech are impoverished, despite all. The principal perceptual correlate of sentence intonation, the fundamental frequency of phonation (Lieberman, 1967), is absent from sinusoidal signals, which imitate only the frequency variation of the formant peaks. As a result of this deficiency, listeners have consistently reported that sinusoidal sentences exhibit noticeably weird patterns of intonation.[2] The perception of relative syllable stress (Fry, 1958; Lehiste & Peterson, 1959; Morton & Jassem, 1965), or of the placement of clause boundaries (Collier & t'Hart, 1975; Lehiste, 1973; Streeter, 1978), each of which is said to follow occasionally from normal intonation, must therefore be supported (if at all) by other means because the anomalous intonation of sinusoidal replicas is quite different from the normal intonation patterns to which these roles are attributed. To the same extent that the fundamental frequency of an utterance also contributes segmental information (about consonant voicing [Summerfield & Haggard, 1977] or vowel identity [House & Fairbanks, 1953], for example), the listener will also be forced to rely on other, alternative sources.

But why do sinusoidal signals create this impression of peculiar intonation in the first place? Prosodic perception is an admittedly complex affair, in which the properties of a single piece of the acoustic stream may affect the recognition of segmental, syllabic, and syntactic structural properties together. In the sinusoidal case, it seems that the pattern of tones imitating only the formant variation inadvertently presents an effective stimulus for perceiving intonation. It is far from obvious why three tones in the frequency range of formants should lead to this impression of vocal pitch, for the acoustic properties corresponding to intonation typically occur several octaves below the lowest formant, and, consequently, below the lowest frequency tone in our three-tone patterns. We undertook the present study to identify the acoustic and perceptual basis for this peculiar concomitant of phonetic perception with sinusoidal signals. The first experiment described here determined which of the likely acoustic sources for the anomalous intonation would in fact be identified as the correlate of sinusoidal intonation. The second experiment tested the salience of the empirically determined acoustic correlate of sinusoidal intonation, the tone reproducing the pattern of the first formant (Tone 1), as a function of its relative amplitude in the three-tone pattern. The third experiment revealed that subjects did not hear the intonation of a sinusoidal sentence as the correlate of Tone 1 when that tone was removed from the sinusoidal sentence pattern. Finally, the fourth experiment that we describe found that the intonation of a four-tone pattern, composed of three sinusoids imitating formant variation and a fourth imitating fundamental frequency variation, was again correlated with the first formant tone and not with the lowest frequency tone of the pattern, complementing the results of the first three studies.

Experiment 1

From the outset, there seemed to be at least three potential causes of the perceptual impression that sinusoidal replicas of natural utterances possess "odd" intonation. First, the apparent speech melody may be the listener's invention, given that the structure of the sinusoidal signal is defective precisely in representing the fundamental frequency of the original utterance. Typically synthetic speech, on the other hand, is generated with a

fundamental frequency pattern as well as a sequence of spectrum envelopes approximating the natural case. In the sinusoidal instance, the listener may fabricate an intonation pattern from the variation in the amplitude envelope of the signal, which is correlated with variation in fundamental frequency in the natural case (Lieberman, 1967), and which also is represented faithfully in sinusoidal replications of natural utterances.

Second, the listener may induce a pitch contour based on whatever changing harmonic relationships exist among the three tones of the sinusoidal pattern. The three tones are not likely to be related harmonically at any given instant, because they follow the computed resonance peaks and not the frequencies of the harmonics of the fundamental closest to the formant centers. Nonetheless, there may be a kind of auditory induction occurring, based on the varying relation of the frequencies of the three simultaneous tones, that produces a time-varying residue heard as the intonation contour. This possibility would be similar to the induction of the missing fundamental (Licklider, 1956; Schouten, 1940).

Third, the listener may use one of the three tones both for segmental information and for intonation information. Although the principal acoustic correlate of sentence intonation is the fundamental frequency, and although the fundamental frequency is present in the speech spectrum at an average of two octaves below the first formant, both psychophysical and electrophysiological evidence suggest that listeners may detect the fundamental frequency of natural utterances by attending to the periodicity of the harmonics of the fundamental in the vicinity of the first formant (Greenberg, 1980). If an extrapolation of those findings is appropriate to the sinusoidal case, we would expect the apparent intonation to be based on the pitch of the tone replicating the first formant of the natural utterance on which it is modeled.

To determine the basis for the apparent intonation of sinusoidal sentences, we performed a test of the apparent similarity of pitch contours, in which subjects judged one member of a pair of tone patterns as more like the speech melody of a sinusoidal sentence. The set of candidate intonation patterns included each of the three tones of the sinusoidal sentence pattern presented individually, a plausible fundamental frequency pattern derived from the amplitude envelope of the sinusoidal sentence, and a tone that reproduced the pattern derived by computing the greatest common divisor of the three tones at intervals of 10 ms throughout the sentence. On each trial, the subject was asked to identify the sentence melody of a three-component sinusoidal sentence presented once, and then to select the single-tone pattern from the two alternatives that was more like the melody of the sentence.

Method

Subjects. Fifteen adults with normal hearing in both ears were recruited by handbill from the combined populations of Barnard and Columbia Colleges. All were native speakers of English, and none had participated in other experiments employing sinusoidal signals. Subjects were paid for their services.

Stimuli. The acoustic materials used in this test consisted of six sinusoidal patterns--one three-tone sentence pattern and five single-tone patterns--produced by the sinewave synthesizer at Haskins Laboratories. This software synthesizer generates sinusoidal patterns defined by parameters of

frequency and amplitude for each tone, updated, in this case, at the rate of 10 ms per parameter frame. The initial synthesis parameters were obtained by analyzing a natural utterance, the sentence "Where were you a year ago?" produced by one of the authors. This utterance was recorded on audiotape in a sound-attenuating chamber and converted to digital records by a VAX 11/780-based pulse-code modulation system using a 5 kHz low-pass filter on input and a sampling rate of 10 kHz. At 10 ms intervals, center-frequency and amplitude values were determined for each of the three lowest formants in this utterance by the analysis technique of linear prediction. In turn, these values were used as sinewave synthesis parameters after correcting the errors that linear prediction is prone to commit. Generally, inappropriate values are easy to identify in the parameter table. They are likeliest to be found when the formant extraction routines are unable to identify any amplitude peaks in the spectrum, for example, when amplitude is low due to consonant closures. Formant patterns are also corrected if the analysis designates an extraneous "formant," which displaces the proper values to the next highest or lowest formant, for example, during rapid spectrum change. A full description of sinusoidal replication of natural speech is provided by Remez, Rubin, and Carrell (1981).

The sentence pattern that was matched to the natural utterance was composed of three time-varying sinusoids. Tone 1 corresponded to the first formant, Tone 2 to the second, and Tone 3 to the third. A Fourier spectrum for a section of the three-tone pattern is shown in Figure 1A. Note that the relative energy in the three tones decreases with increasing frequency, imitating the natural case, but that the broadband formant and harmonic structure common to voiced speech is not present. The five alternative single-tone patterns that were used to compose the pairs of alternatives were: Tone 1, Tone 2, or Tone 3, each a component of the sentence pattern that the subject heard at the beginning of each trial; a plausible fundamental frequency pattern (PFO) computed from the amplitude envelope; and the pattern comprising the values of the greatest common divisors (GCD) of the three concurrently varying tones in the replica of the natural utterance, computed for each 10 ms frame of the sinusoidal synthesis parameters. Each of the single tone alternatives was produced with equal average power. The PFO was derived by modulating the frequency of a 100 Hz tone to follow the changes in amplitude of the waveform of the sinusoidal sentence. The maximum range of this tone was 20 Hz, and the maximum rate of frequency change was 1 Hz/10 ms. Finally, the frequency values for synthesizing the "missing fundamental" tone were determined by computing the integer, for each synthesis frame, of greatest value that served as a divisor for each tone frequency, with no more than a 2% remainder. The average frequency value of this plausible missing fundamental tone was 92 Hz, well within the fundamental range of the talker producing the original utterance from which these six tonal patterns were derived. The amplitude values of this tone were matched for each 10 ms frame to the amplitude values of Tone 1. A graphic representation of each of the five single-tone patterns is presented in Figure 1B.

The synthesized test materials were converted from digital records to analog signals, recorded on audiotape at Haskins Laboratories, and were presented to listeners in the Perception Laboratory of the Department of Psychology, Barnard College, by playback of the audiotape. Average signal levels were set at 72 dB SPL. Stimuli were delivered binaurally in an acoustically shielded room over Telephonics TDH-39 headsets.

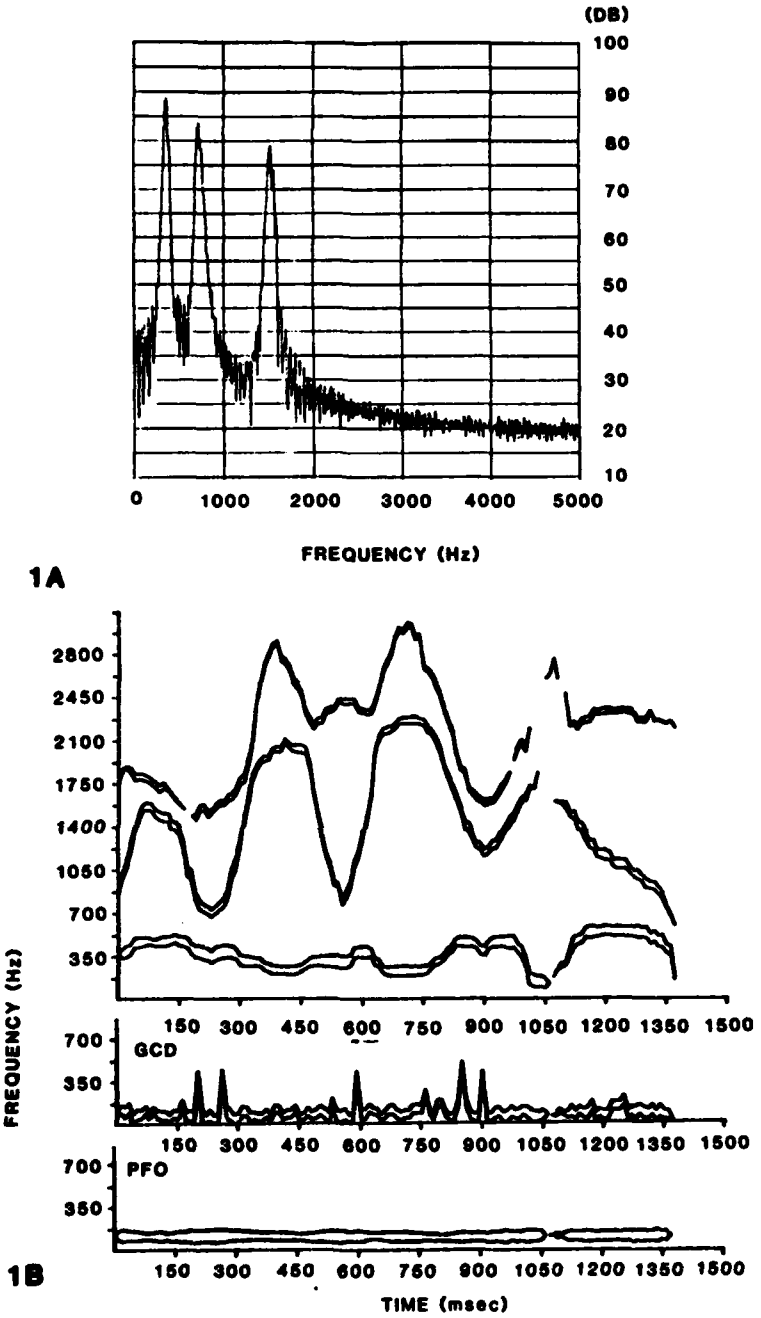


Figure 1. (A) This panel is the Fourier spectrum of a representative section through the three-tone pattern replicating the sentence "Where were you a year ago?" (B) The five tone patterns used as stimuli in Experiment 1. Top panel: The three-tone pattern replicating the first three formant center-frequency values of the sentence "Where were you a year ago?" Middle panel: The pattern composed of the greatest common divisors (GCD) of the three tones in the sentence replica, computed for each 10-ms synthesis frame. Bottom panel: A plausible fundamental frequency pattern (PFO), computed from the amplitude envelope of the sentence pattern. In all cases, variation in thickness represents amplitude variation.

Procedure. Listeners were instructed that the experiment was examining the identifiability of vocal pitch, the tune-like quality, of synthetic sentences. To illustrate the independence of phonetic structure and sentence melody, the experimenter sang the phrases "My Country 'Tis of Thee" and "I Could Have Danced All Night" with the original melodies and with the melodies interposed. When subjects acknowledged they could determine the melody of a sentence regardless of the words, they were instructed a) to attend on each test trial to the pitch changes of the sinusoidal sentence, b) to identify the pattern, and c) to select which of the two patterns more closely resembled the pitch of the sentence. Subjects recorded their choices in specially prepared response booklets.

Each trial had the same format. First, the sinusoidal sentence "Where were you a year ago?" was presented once. Then, one of the five single-tone patterns was presented. Finally, a second single-tone pattern was presented. There were ten different comparisons among the five different single-tone alternatives. Counterbalanced for order, each subject judged each different comparison ten times. Each sinusoidal pattern was approximately 1400 ms in duration; the interval between items within a trial was 1 s; and, the silent interval between trials was 3 s.

Results and Discussion

An analysis of variance was used to identify the differences among the means of subjects' performance in the differential similarity test. Irrespective of the order of alternatives within a trial, there were ten different trial types comparing tonal alternatives: Tone 1 vs. Tone 2; Tone 1 vs. Tone 3; Tone 1 vs. PFO; Tone 1 vs. GCD; Tone 2 vs. Tone 3; Tone 2 vs. PFO; Tone 2 vs. GCD; Tone 3 vs. PFO; Tone 3 vs. GCD, and PFO vs. GCD. For each type of trial, a signed value indexing the preference for one alternative or the other was computed by taking the difference of the number of trials (out of ten) on which the subject selected the first alternative versus the second. (The order of alternatives used to determine the sign of the difference was the order of the alternatives given directly above.) Note that if the subject had no consistent preference within a trial type, the index value approached 0, while a consistent preference approached (+ or -) 10. The one-way analysis of variance revealed a significant difference among the means of the similarity scores on different trial types, $F(9,126) = 11.8$, $p < .001$. Scheffé post hoc means tests showed that Tone 1 was preferred to every alternative with which it was compared, but that in trials excluding Tone 1 the greatest performance difference was not significant. Histograms of the group data are shown in Figure 2. The figure represents the proportion of trials on which each alternative in each comparison type was selected. From this figure it seems clear that the tone replicating the first formant is chosen as the sentence pitch in any comparison that includes it (2A), and that in every other case the choice of tone is equivocal (2B).

The outcome supports a few conclusions about the cause of the odd intonation of sinusoidal sentences. It seems that the tone that replicates the first formant of the natural utterance is put to two uses, perceptually, by listeners. Although it seems to provide segmental information about consonants and vowels, as we expected, it also serves as the acoustic correlate of sentence pitch, a function usually attributed to the fundamental frequency of phonation. This outcome seems surprising because Tone 1 in sinusoidal sentences is typically one and one-half octaves higher than the fundamental, as

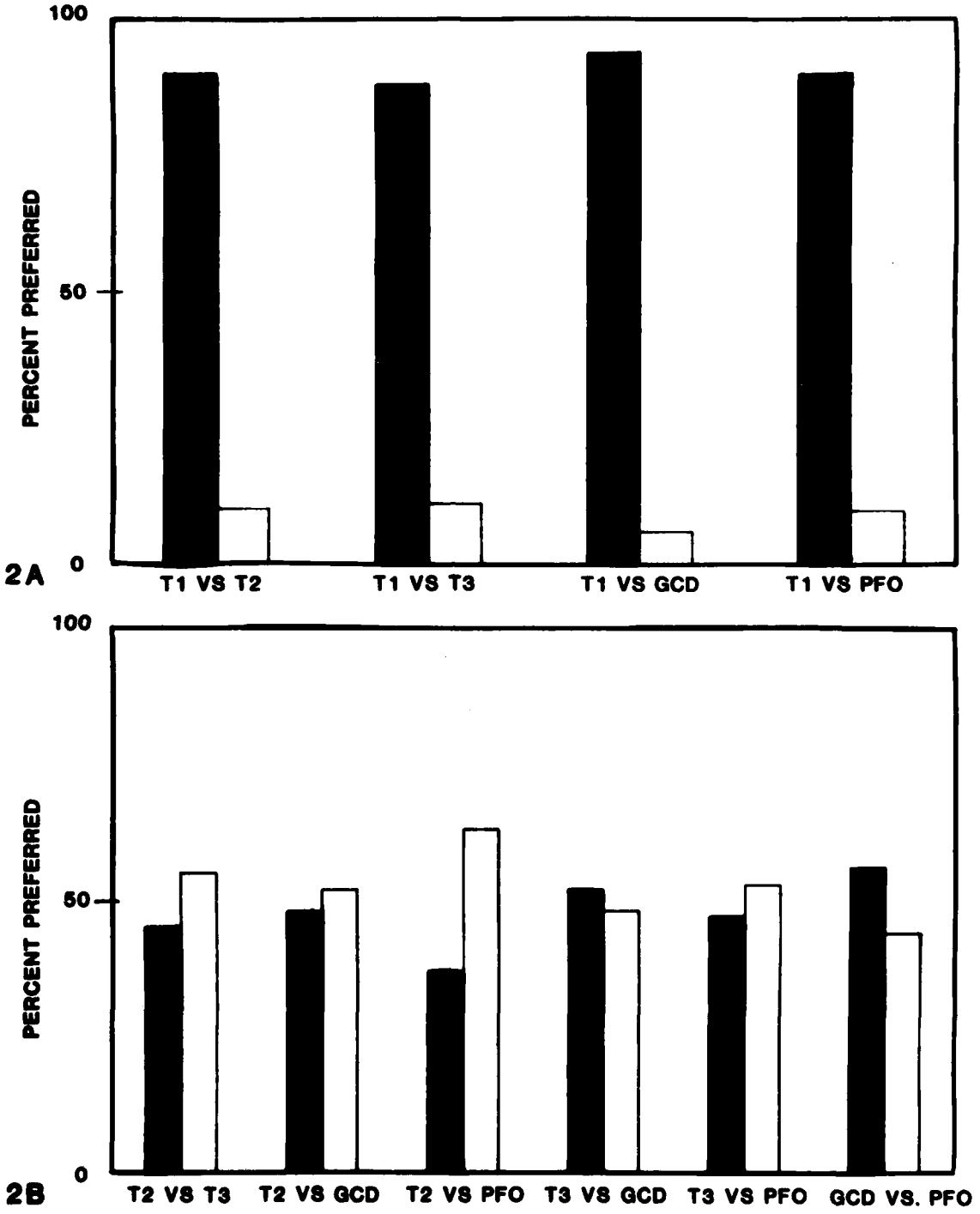


Figure 2. Differential similarity data from Experiment 1. (A) Comparisons in which Tone 1 was an alternative. (B) Remaining comparisons.

is the first formant in natural utterances. Moreover, Tone 1 would be quite beyond the comfortable phonatory range of adults capable of producing the associated formant frequencies. The perceptual preference for Tone 1 as the intonation of the sentences is not to be expected, therefore, if perception is based primarily on the listener's knowledge of the normative articulatory abilities of talkers. Although some evidence implies that the variation in fundamental frequency in natural speech is correlated with formant pattern (House & Fairbanks, 1953; Lehiste & Peterson, 1961), no current proposal also suggests that the perceiver uses the first formant frequency variation as information both for intonation and for segment identification. This, however, seems to have occurred in the case of sinusoids.

Research on the phenomenon of the dominance region (for example, Plomp, 1967; Ritsma, 1967; see also Greenberg, 1980) may begin to explain this result. These studies established that the impression of pitch corresponds to the shared fundamental period of the third through fifth harmonics, and not to the periodicity of excitation occurring in the lower or higher frequencies. In the nonspeech case (Plomp, 1967), listeners judged the apparent pitch of complex signals composed simultaneously of two different harmonic series. Each series presumably could have led to the impression of a different pitch, but the series falling within the "dominant" region in fact determined the pitch. In the speech case, Greenberg (1980) recorded evoked potentials to synthetic vowels in human subjects. He found that the auditory representation of fundamental frequency was strongest when the first formant occurred within the dominant region. If the impression of pitch is obtained from analysis of this band in the auditory representation, then the implication of this work is clear: A person listening to speech normally uses the region of the spectrum associated with the first formant to obtain periodicity information as well as to detect the frequency of the first formant itself. Ordinarily, the periodicity of the stimulus in this region and the frequency of the first formant will differ, although in the present case they happen to be identical.

We cannot be sure, however, that Tone 1 is selected for its prosodic role for any reason other than it is the loudest tone in the three-tone complex. Recall that the parameters specified for each time-varying sinusoid in the replication of the natural utterance include a formant center frequency and a formant amplitude specification, both derived by linear prediction analysis of the speech waveform. Because the first formant in natural speech commonly has the greatest energy and each higher formant less energy, this spectrum envelope rolloff is therefore preserved in the sinusoidal imitation. To identify the relation between the selection of Tone 1 as the pitch contour of the sinusoidal sentence and its relatively great acoustic power, we performed Experiment 2. In addition, we attempted to test the generality of our finding by using a new sentence.

Experiment 2

In this portion of our study, we varied the relative amplitudes of the three tones composing the sinusoidal sentence, and again employed a test of differential similarity to determine the alternative most similar to the intonation of the sinusoidal sentence. If, in Experiment 1, Tone 1 was judged most similar in pitch pattern to the intonation of the sinusoidal sentence merely because Tone 1 had the greatest energy of the three components of the sentence pattern, then this should not recur when the relative amplitude differences of the three tones are eliminated, or reversed. On the other hand, if Tone 1 is the stimulus for intonation because it occurs within the

dominance region, then we should not expect amplitude variation to change the differential similarity performance, as long as Tone 1 is detectable (Ritsma, 1967). Experiment 2, therefore, estimated the effects on apparent intonation of equating the amplitudes, and of inverting the order of amplitudes, among the tones of a three-component replica of a natural utterance. In addition, we also used a different sentence in order to identify any effects that may have been particular to the phonetic properties of the sentence used in the first experiment.

Method

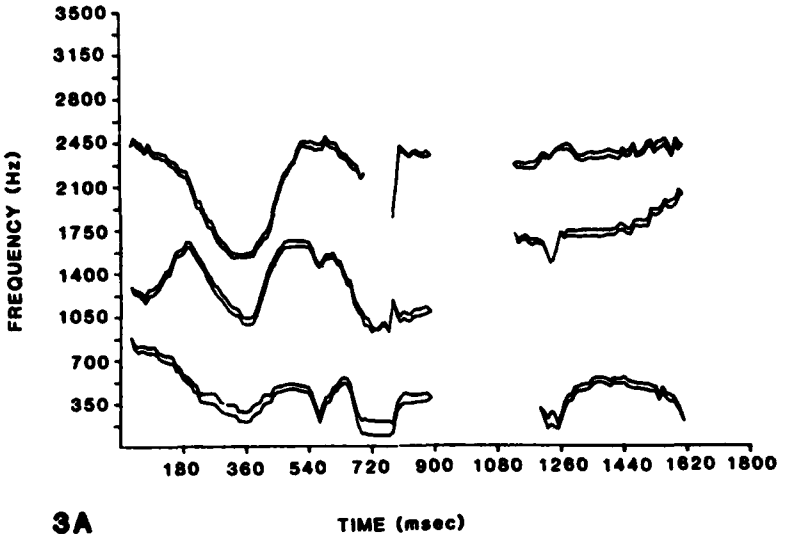
Subjects. Fourteen listeners were again drawn from the local population of audiologically normal undergraduates. None had been tested previously in studies of sinusoidal synthesis. Subjects received pay for participating.

Stimuli. A three-tone replica was prepared for the sentence "I read a book today," according to the procedure described in Experiment 1. Two versions were subsequently made from this replica. In the first, the tone amplitudes were set equal; in the second, the amplitude order was the inverse of the natural case, with Tone 3 possessing the greatest power and Tone 1 the least. Figure 3A shows the pattern of three tones composing the sentences. Figures 3B and 3C show Fourier spectra of sections of the equal (flat) amplitude and uptilted amplitude versions of this sentence.

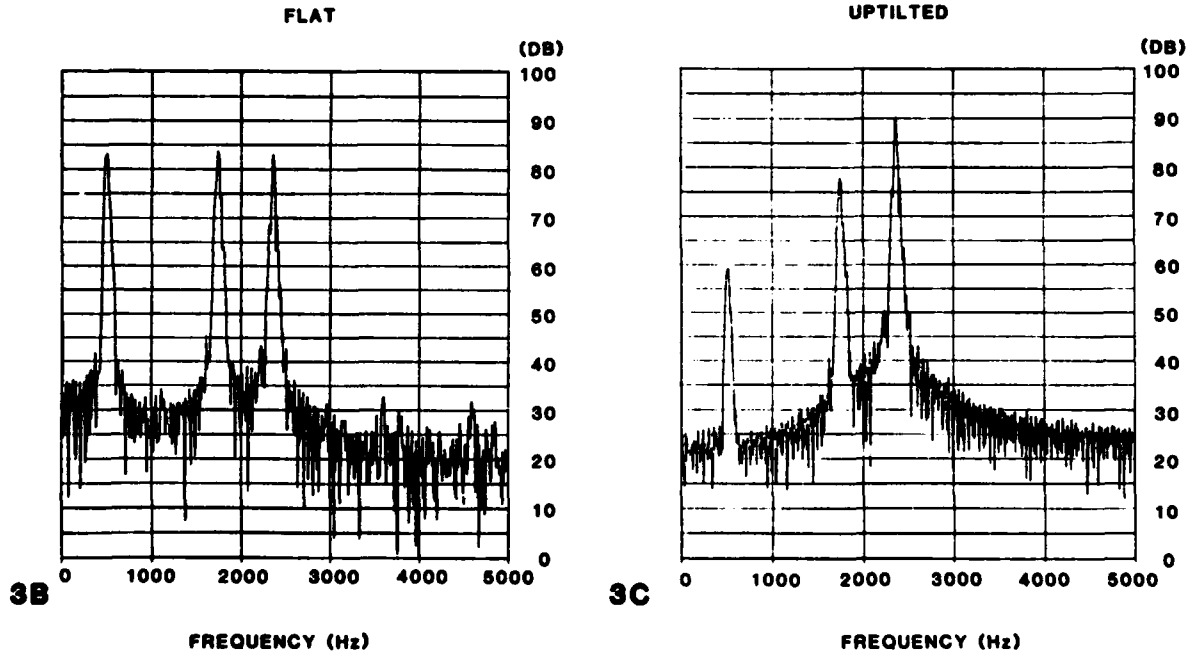
The single-tone alternative patterns to be compared with the apparent intonation of the sinusoidal sentence on each trial consisted this time simply of each of the three individual tones composing the sentence. The single-tone alternatives were prepared as in Experiment 1, with equal average power. On each trial, the subject heard one of the two versions of the sentence, with the flat or the uptilted spectrum, followed by two of the three alternative tone patterns.

Procedure. Each trial began with a single presentation of the sinusoidal sentence "I read a book today," in either the flat or the uptilted spectrum version. Two single-tone alternatives followed, from which the subject selected the better match to the apparent intonation of the sentence. Collapsing over the counterbalancing of order for each pair of alternatives, there were three different types of trials: the comparison of Tones 1 and 2, Tones 1 and 3, and Tones 2 and 3. Each of these was presented twenty times, ten times in each order. In addition, twelve trials were interspersed in the test order in which a normal spectrum relationship occurred among the tones of the sentence, although the overall power was greatly reduced. The only alternative tonal intonation patterns presented for this quiet, normal-amplitude rolloff sentence were Tones 1 and 2. The data from this condition served as a converging check on the outcome of the first experiment.

One hundred and thirty-two trials were presented in this test. On each trial, the subject first identified the intonation of the sinusoidal sentence presented and then selected the more similar of the two lagging alternative tone patterns. The choice was recorded in pencil on a specially prepared response booklet. There were intervals of 1 s between items within a trial, 3 s between trials, and 8 s following every twelfth trial.



3A



3B 3C

Figure 3. (A) The three-tone pattern replicating the sentence "I read a book today." (B) Fourier analysis of a section through the flat spectrum version of the sentence (equal energy in each tone). (C) Fourier analysis of a section through the Uptilted spectrum version. Compare to Figure 1A.

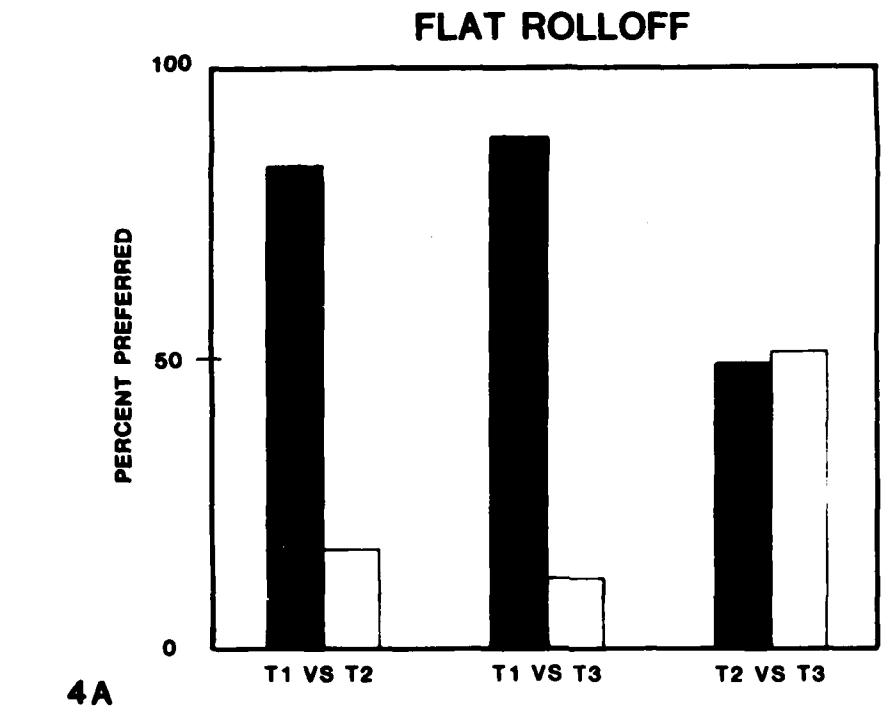
Results and Discussion

The judgments were handled in a manner analogous to Experiment 1. Signed preference scores were determined for the three comparisons of the flat and uptilted sentence conditions. For each comparison, the difference was computed between the number of trials on which the first alternative was chosen and the number of trials on which the second was chosen. In the computation of the different scores, the alternatives were compared in this order: Tone 1 vs. Tone 2, Tone 1 vs. Tone 3, Tone 2 vs. Tone 3. A two-way repeated measures analysis of variance, with the factors SENTENCE (flat vs. uptilt) and COMPARISONS (Tones 1 vs. 2, 1 vs. 3, and 2 vs. 3) was used to determine whether there was an effect of tone amplitude in the sentence on the perception of intonation. The data from the quiet, normal-amplitude trials, in which Tone 1 was the clear preference, were omitted from this analysis.

The group data are shown in Figure 4. It is obvious from that figure that Tone 1 retains its preferred status. This is confirmed by the analysis of variance. There was a main effect of sentence type, indicating that the preference scores were more consistent for the flat than for the uptilted sentences, $F(1, 13) = 9.5$, $p < .01$; in addition, there was a main effect of trial type, $F(2, 26) = 10.1$, $p < .001$, with Tone 1 preferred to each of the two pairs in which it occurred, and no consistent preference between Tones 2 and 3. The interaction term was not significant, $F(2, 26) = 0.6$, $p > .1$, indicating that the subjects preferred Tone 1 as the best match for sentence intonation regardless of the spectrum manipulation.

This experiment supports the conclusion of our first experiment on sinusoidal intonation. It seems that the functions of Tone 1 include both the segmental use typically associated with the first formant that it replicates, and the use typically identified with the fundamental frequency of phonation in natural speech. The durability of the listener's reliance on Tone 1 for intonation information is noteworthy, especially considering the inversion of the order of relative amplitudes among the tonal components of the signal. It suggests that the dual use of Tone 1 in sinusoidal sentences is brought about by virtue of its occurrence within the dominance region, and not because it is the component with greatest power. Periodicity within this frequency band, including instances of relatively low power, evidently determines the pitch pattern of the perceived sentence. It seems, then, that Tone 1 is concurrently represented as an amplitude peak in the spectrum, which provides information about segmental phonetic properties of the utterance, and also as a periodic spectrum element that determines the apparent pitch of the tonal complex. Ordinarily, in speech, the frequencies occurring within this region are harmonics of the fundamental frequency of phonation. However, in this anomalous case of formant center frequencies without harmonic excitation, there is no stimulation, periodic or otherwise, in the range of a talker's fundamental, and therefore no harmonics in the dominance region. There is, simply, the time-varying frequency of the tone following the formant, which is treated as the stimulus for pitch by default, regardless of its amplitude relative to the other components.

To conclude that the intonation of a sinusoidal replica is the correlate of Tone 1, and that this is attributable primarily to the occurrence of this time-varying periodic tone within the dominance region of the auditory system, we must establish that listeners reject Tone 1 as the best match of sinusoidal sentence intonation when the sentence does not include that tone. In other



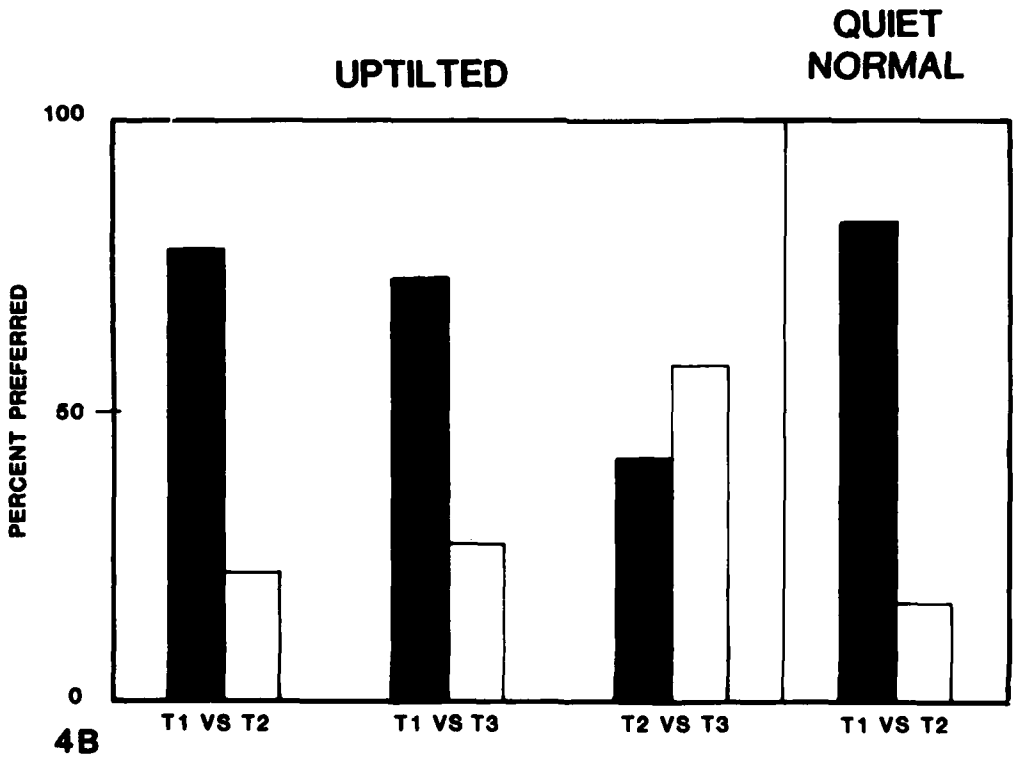


Figure 4. Results of Experiment 2. (A) Differential similarity data for flat rolloff condition. (B) Differential similarity data for the uptilted and quiet-normal spectrum conditions.

words, if a two-tone pattern, including only Tones 2 and 3, is presented in the same paradigm as in Experiments 1 and 2, listeners should not report that Tone 1 matches the intonation of this pattern. Were they to persist in identifying Tone 1 as the intonation pattern, we would be forced to conclude that the phenomenon of sinusoidal intonation is less a matter of the ordinary perception of extraordinary signals, as we have alleged, and actually is a matter of special induction of ad hoc attributes of an unfamiliar stimulus. Experiment 3 was performed to test whether Tone 1 is identified as the correlate of intonation for patterns that do not contain it.

Experiment 3

At this point, the evidence shows that the tone following the frequency variation of the first formant is the correlate of the intonation of sinusoidal sentences. In all cases, Tone 1, corresponding to the track of the first formant, was judged more like the sentence intonation than any other candidate. Our conclusion has emphasized the listener's tendency to identify the periodicity of the stimulus by attending to the dominance region, and to perceive pitch from the representation of stimulus frequency within that region. Independent evidence from studies of nonspeech tones and vowels supports the general conclusion that frequency in the dominance region causes apparent pitch, even for natural speech. Hence, the explanation of sinusoidal intonation that we offer is that these atypical stimuli are evaluated perceptually in essentially the same manner as are nonspeech tonal complexes and speech sounds.

However, simply because subjects choose Tone 1 consistently as the best match to apparent pitch does not mean that Tone 1 is causing the pitch percept. To support this characterization of the perception of sinusoidal intonation, we must determine that subjects do not select Tone 1 when it is absent from the tonal sentence. If subjects select Tone 1 as the match to intonation only when it is present in the sentence, then we would have reasonable grounds to support our stimulus-based hypothesis of the phenomenon. Otherwise, if subjects continue to prefer Tone 1 to other candidate tones when that tone is omitted from the sinusoidal sentence, then we would necessarily conclude that intonation occurred through a form of auditory induction, however similar this induced pattern would be to the pitch contour of Tone 1. Experiment 3 evaluated this possibility by presenting a test of differential similarity in which the sentences to be matched contained either the three tones corresponding to the first three formants or merely the tones corresponding to the second and third formants, omitting the first.

Method

Subjects. Twenty-one listeners were selected as before from the student population of Barnard and Columbia Colleges. None had participated previously in experiments of this nature. They were paid for their participation.

Stimuli. The three-tone sinusoidal replicas of the sentence "I read a book today" prepared in Experiment 2 provided the basis for all stimuli in this test. Three versions of the sentence were used. The first was the uptilted amplitude replica, in which the tone amplitudes were the inverse of the natural case of formants. Tone 1 had the least power, and Tone 3 the most. The second sentence was the pattern consisting only of Tones 2 and 3 of this replica. This two-tone pattern was equated informally, by the authors,

for loudness equal to the three-tone pattern. Note that Tone 1 is omitted from this pattern. The third sentence was the three-tone replica, preserving the natural amplitude relations among the tones but presented at low power, again to serve as a check on the outcome. The three single-tone patterns from Experiment 2 were used as alternative pitch contours in this test of differential similarity.

Procedure. Listeners were instructed to identify the sentence melody of the sinusoidal sentence presented first on each trial, and then to select the better match to that sentence melody from the two lagging single-tone alternatives. Subjects were urged not to omit judgments. The choices were scored in pencil in specially prepared response booklets.

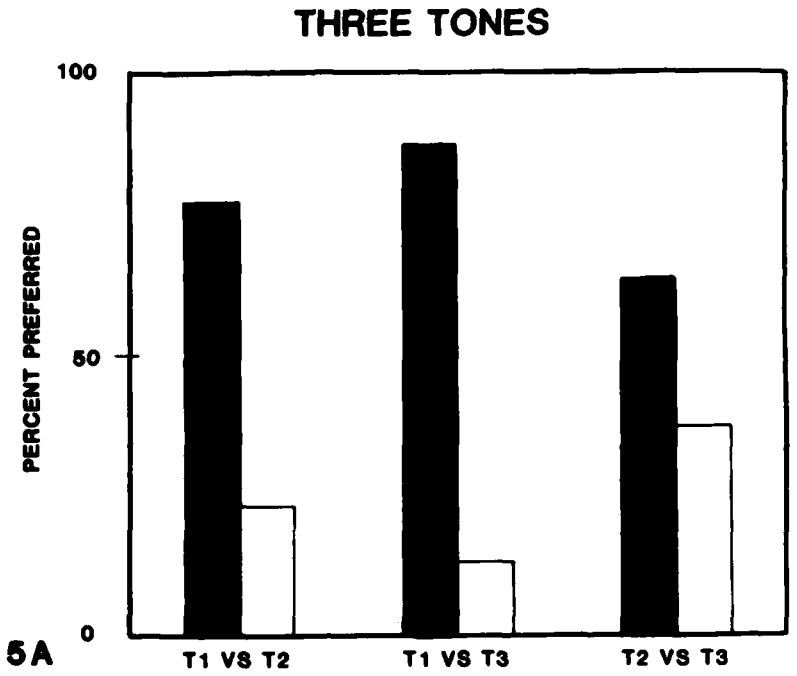
There were three different combinations of alternatives, counterbalanced for order of presentation: Tone 1 vs. Tone 2, Tone 1 vs. Tone 3, and Tone 2 vs. Tone 3. Each trial type was presented twenty times in random order with each of the two sentence versions, Three-Tones and Tones 2 and 3. A third sentence, Normal-Quiet, occurred twelve times in this test paired only with Tone 1 vs. Tone 2 alternatives. The test, then, consisted of 132 trials. Within a trial, the three patterns were separated by intervals of 1 s. Trials were separated by 3 s, with 8 s between blocks of 12 trials.

Results and Discussion

The results of the similarity judgments are shown in Figure 5. It is clear that subjects once again selected Tone 1 when it occurred as a component of the sentence. In the case of the sentence containing only Tones 2 and 3, however, subjects instead preferred Tone 2 to Tone 1 as the best match for the sentence intonation. This outcome corresponds to a highly significant interaction term in the analysis of variance, $F(2, 40) = 52.4$, $p < .001$. Subjects also preferred Tone 2 when it was pitted against Tone 3 in the context of the two-tone sentences. Overall, subjects reported that sentence pitch was matched best by Tone 1 only when that tone was a component of the sentence.

This third experiment is encouraging with respect to the hypothesis we offered about sinusoidal intonation. Subjects appear to be treating these anomalous signals in a manner similar to speech signals. It is as if the segmental information is obtained from the formant-like frequency variation of the tones, and intonational information is provided by the periodicity within the dominance region. This occurs despite the congruence of these two kinds of information in the pattern of frequency variation of Tone 1.

However, to establish the appropriateness of this application of the dominance region notion, we must perform one final test. This is necessitated by the kind of evidence we have obtained so far on the predominance of Tone 1 in producing the apparent intonation. Although our experiments have shown that listeners consistently judge this tonal component to be most like the sentence melody of a sinusoidal utterance, we have not separated two aspects of this tone within the three-tone pattern that composes a sentence. In the three tests that we report, the tone corresponding to the first formant has been both the tone within the dominance region and the tone with the lowest frequency, overall, in the three-tone complex. Because of this fact, we cannot distinguish empirically between the dominance region hypothesis and a lowest-frequency component hypothesis. To do so requires a test in which the subjects evaluate a sentence that contains tonal components falling in the



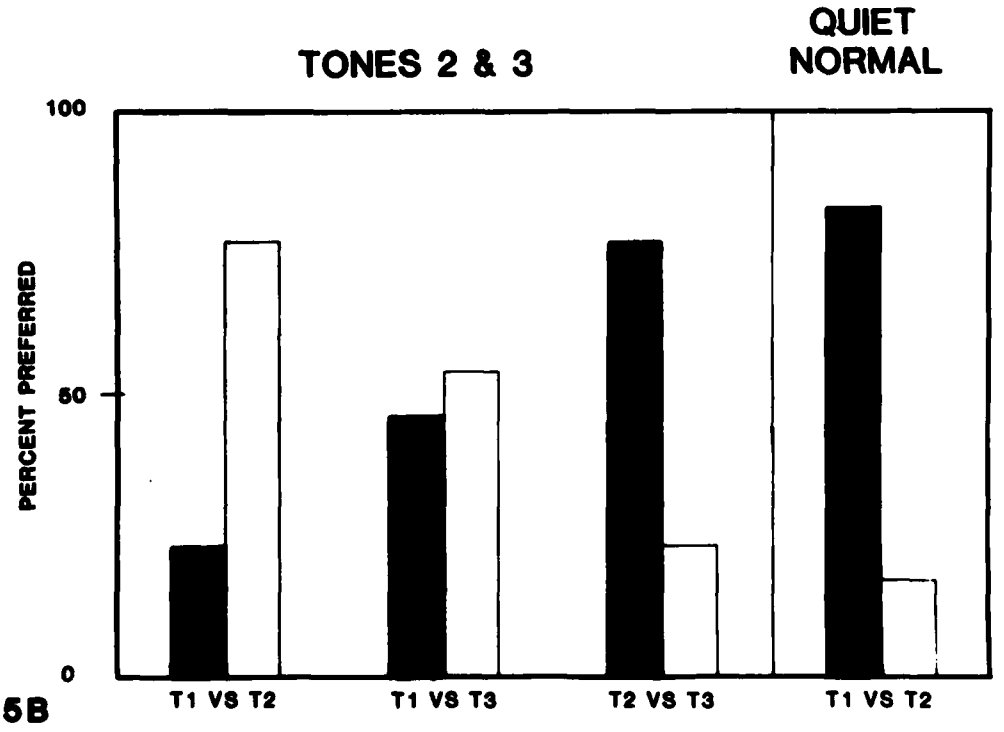


Figure 5. Results of Experiment 3. (A) Differential similarity data for the three-tone sentence with flat rolloff. (B) Differential similarity data from the two-tone and quiet-normal conditions.

dominance region, below the dominance region (with frequencies < 400 Hz) and above the dominance region (with frequencies > 1000 Hz). We can predict the outcome based on Experiments 1-3: When subjects listen to such a sentence, they should either attend to the tone within the critical frequency range for perceiving intonation, which would encourage the dominance region explanation that we have proposed; or, they should prefer the lowest frequency tone, which would falsify the dominance region hypothesis, though in a manner consistent with the findings that we have noted throughout this investigation. This test is the topic of Experiment 4.

Experiment 4

The original rationale for the dominance region was that the auditory system gets the stimulus for pitch where the harmonics are resolved the best. At this juncture, we have shown the superiority of Tone 1 (corresponding to the first formant) compared to simultaneously occurring tones with higher frequencies. Additionally, the dominance region hypothesis predicts that listeners should also reject tones falling below the dominance region. To perform this test of the claim, we returned to the natural utterance of our familiar test sentence, and analyzed its fundamental frequency pattern. From this analysis, a new set of sinewave synthesis parameters was created to form a tone with a pattern of frequency variation matching the natural fundamental frequency contour. These values were used in combination with the three-tone replica to generate a four-tone sentence, comprising a "fundamental frequency" tone and the three "formant" tones, as well as the additional single tone alternative to use in the similarity test format.

In the four-tone sentence that subjects evaluated, the tone matching the fundamental frequency contour falls below the dominance region. If the likeness of the first formant tone to the apparent sinusoidal intonation is based on its occurrence within the critical frequency range, then we may expect listeners to reject the fundamental frequency tone no less consistently than they have rejected the second and third formant tones in Experiments 1, 2, and 3. In other words, a tone representing a fundamental frequency pattern from a natural utterance should ironically not provide information for sentence melody in this case, despite the naturalness of its pattern of variation and the appropriateness of its occurrence in the normal frequency range of the fundamental frequency.

Method

Subjects. Twenty-four listeners participated in this study. They each reported a normal history of speech and hearing function, and had not previously been introduced to synthetic speech or sinewave materials. Our subjects were student volunteers who received course credit in exchange for taking this brief test.

Stimuli. The sentence presented to subjects in this test was composed of four tones: Tone 0 corresponding to the fundamental frequency (commonly termed F0) and overall amplitude of the original natural utterance of "I read a book today," on which we patterned the sinewave sentences reported in the previous two experiments; and Tone 1, Tone 2, and Tone 3, each corresponding to the pattern of center-frequency and amplitude variation of the first three formants. The values of the fundamental of the natural utterance were obtained by employing the cepstral method of pitch extraction on the sampled da-

ta, and were converted to sinewave synthesis parameters by including amplitude values varying in imitation of the overall energy of the natural utterance. The pattern of frequency variation of Tone 0 is shown in Figure 6A. The four-tone pattern formed by combining Tone 0 with the three tones that replicate formant variation preserved the natural spectral amplitude rolloff, as shown in Figure 6B.

The test stimuli also included the four sinusoidal components realized as single-tone patterns, to be used as alternative pitch stimuli in the similarity test. Each of the tones was resynthesized in isolation and the four were equated for loudness.

As before, the test sequence was recorded on audiotape, and presented to listeners via playback and calibrated headsets. An average listening level of 72 dB SPL was used.

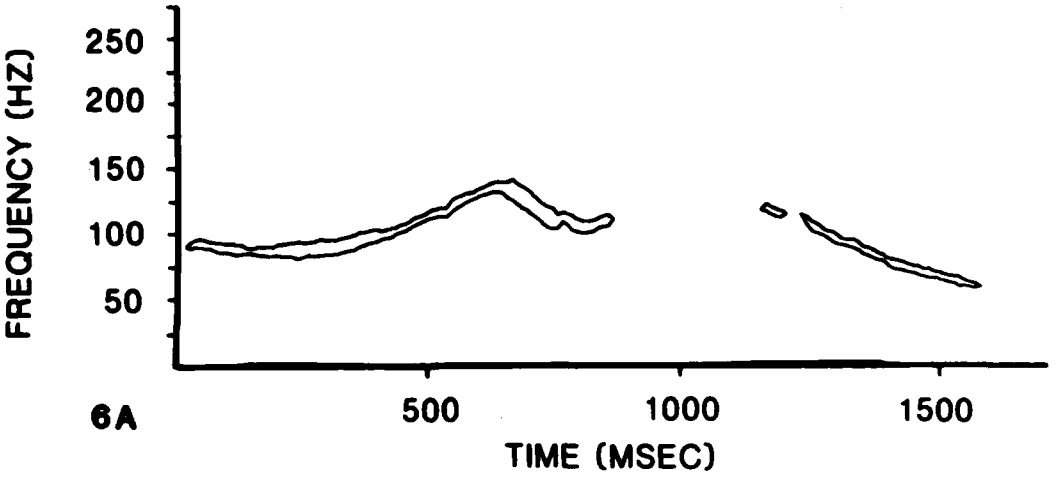
Procedure. A test of apparent similarity was again used in this experiment. Each trial consisted of three sinusoidal patterns: first, the four-tone sentence pattern, followed by two single-tone patterns. There were six different trial types, exhausting the possible comparisons among the four single-tone candidates: Tone 0 vs. Tone 1, Tone 0 vs. Tone 2, Tone 0 vs. Tone 3, Tone 1 vs. Tone 2, Tone 1 vs. Tone 3, and Tone 2 vs. Tone 3. Each was presented in two orders to counterbalance the occurrence of alternatives. Altogether, the test consisted of the six trial types presented 14 times each, including counterbalancing, composing a sequence of 84 trials.

On each trial, subjects were instructed to identify the sentence melody of the first sinusoidal pattern, and then to select the better match of the two lagging alternative patterns. Omissions were discouraged. The judgments were reported with pencil and paper using specially prepared booklets.

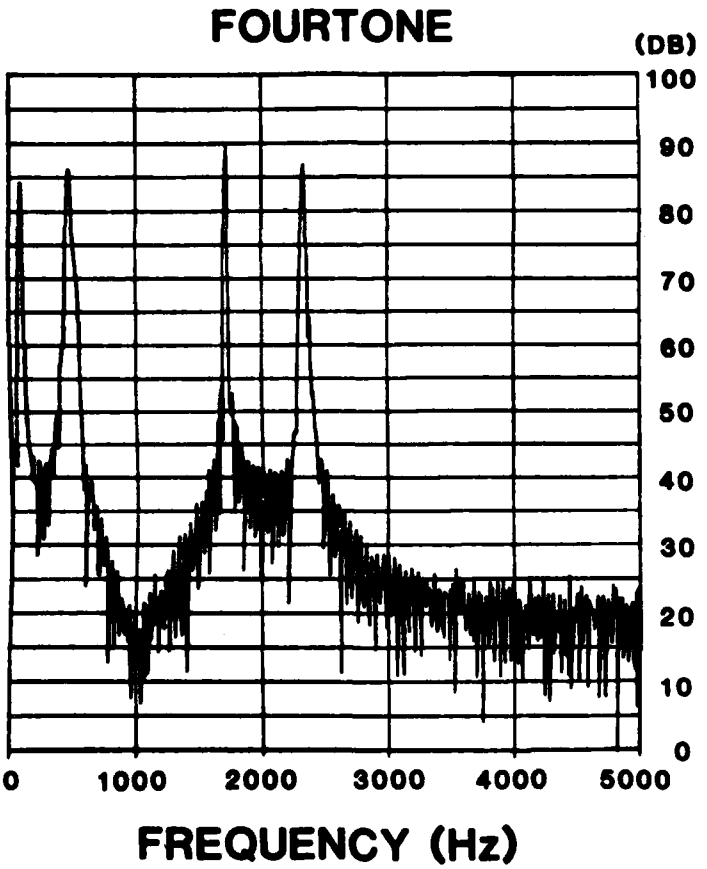
Results and Discussion

The histograms in Figure 7 describe the results of the similarity test. Tone 1, corresponding to the first formant, was once again preferred to every other candidate tone. Tone 2 was judged more similar to the intonation pattern than was Tone 3, an unanticipated effect. And, most critically, subjects rejected Tone 0 consistently when it was an alternative paired with Tone 1, indicating that the impression of sentence melody is stable. These results were confirmed in the analysis of variance of similarity scores, $F(5, 155) = 40.1$, $p < .001$, and by Scheffé post hoc means tests.

The pattern of results of Experiment 4 clearly confirms the appropriateness of the dominance region hypothesis for the phenomenon of sinusoidal sentence intonation. In fact, the congruence of segmental and intonational information in the sinusoidal case of Tone 1 permits us to support a proposal about auditory analysis of natural speech: Fundamental periodicity is represented in the auditory system based on harmonics detected ithin the dominance region and not on attention to the fundamental itself. Because Tone 1 occurs within the range of this normal region for detecting periodicity in the waveform, it seems to be treated as the principal stimulus for pitch perception.



6A



6B

Figure 6. (A) The frequency pattern of Tone 0, which reproduced the fundamental frequency pattern of the natural utterance "I read a book today" in sinusoidal form. (B) A representative section through the four-tone sentence pattern.

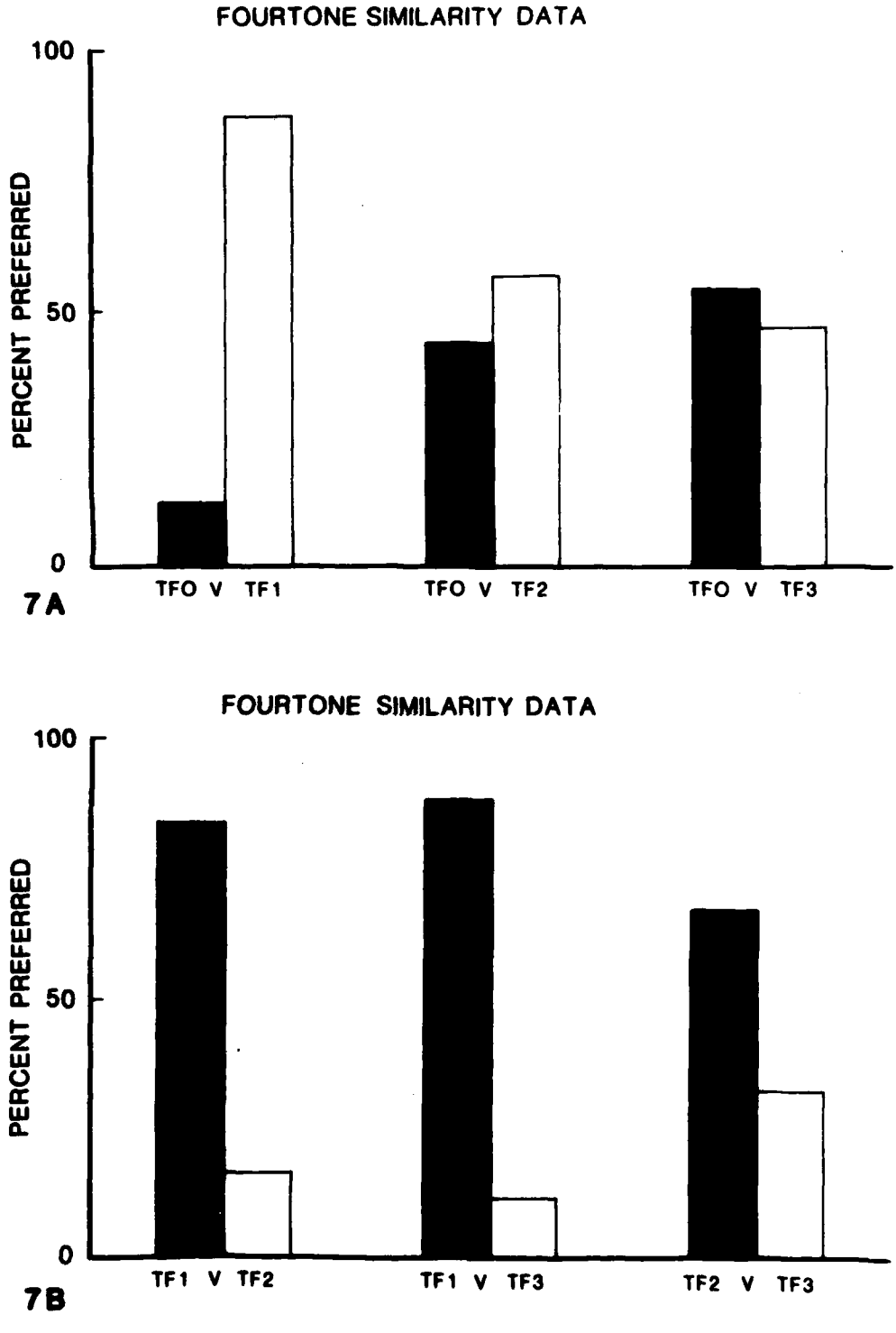


Figure 7. Results of Experiment 4. (A) Differential similarity data for comparisons involving Tone 0. (B) Differential similarity data for remaining comparisons.

General Discussion

Prosody is a perceptual dimension of utterances that is not caused by variation in any single physical dimension of the acoustic signal. The listener is likely to treat the duration, amplitude, and fundamental period of portions of the speech signal as changes in the rhythm, meter, and organization of the linguistic utterances that perception defines. One aspect of prosody is intonation, or sentence melody. The problem for the theorist is to identify the relations among the quite dissimilar physical ingredients that produce impressions of intonation in some cases, but create impressions of duration, or loudness, or lexical stress, or perhaps syntactic constituent boundaries, in others. In addition to the effects of these physical variables, perception of intonation has been viewed as a process that refers to linguistic knowledge, because judgments of intonation often reflect lexical properties (Lieberman, 1965; but see Lea, 1979).

Given the intricate interplay of physical and perceptual components in prosodic perception, it seem anticlimactic to assert that the perception of intonation is based principally on fundamental frequency, in some instances necessarily so (Abramson, 1972). However, intonation is potentially determined from integrated energy or from frequency variation in the third and fifth formants in whispered sentences (Meyer-Eppler, 1957), which lack contours of fundamental frequency. As such, the whispered utterance is the most reasonable precedent for sinusoidal sentence perception. A sinusoidal replica also lacks a fundamental frequency of excitation common to its tonal components, and therefore we might have expected it to be treated in a manner similar to that of a whispered sentence. Instead, we found consistent perceptual reliance on the portion of the signal within the dominance region as the primary ingredient to intonation, much as occurs for normal utterances.

We cannot yet define a principle by which intonation is variously derived from the fundamental, or the amplitude envelope, or the higher formant frequency changes. Because our exploratory studies probed this phenomenon at the sentence level, neither have we determined the extents of the likely influence of duration, amplitude, and relative frequency change, on the one hand, or of lexical access, constituent structure, and the encoding of intonation in memory, on the other. Each of these factors may be suspected of moderating the effect on fundamental frequency. Even if these other influences are slight, we may nevertheless expect intonation to differ from the fundamental frequency pattern (Hadding-Koch & Studdert-Kennedy, 1964). With these cautions in mind, we propose that our investigation describes the perceptual registration of the strongest influence on intonation, the fundamental frequency.

The studies reported here lead us to conclude that speech signals are analyzed for fundamental frequency in the dominance region, coincidentally, the region of the first formant, as Greenberg (1980) hypothesized on the basis of studies of the strength of periodicity in auditory evoked potentials with synthetic vowels. It is somewhat ironic that sinusoidal signals, clearly unnatural in vocal timbre, provide evidence on this question. But, if the auditory system ordinarily detects periodicity from the harmonics in the dominance region, then when it fails to find harmonics it seems nevertheless to represent the pitch of a complex signal by its period in this region. A sinusoidal sentence is a kind of exceptional stimulus that tests the rule, and confirms it.

Is the intonation of sinusoidal sentences the result of periodic acoustic structure subsequently transformed by duration and loudness (or by segmental and morphological structure)? If sinusoidal signals and natural speech are analyzed in a common manner, as we claim, then we may certainly expect sinusoidal intonation to be affected by acoustic and linguistic properties besides frequency of the tone in the critical range. For the present, though, the evidence suggests that the primary correlate of sinusoidal intonation is the tone that reproduces the frequency variation of the first formant. And, while this outcome is revealing about the perception of natural speech, it also supports the contention that sinusoidal replicas of utterances are perceived like ordinary phonetic signals.

References

Abramson, A. S. (1972). Tonal experiments with whispered Thai. In A. Valdman (Ed.), Papers in linguistics and phonetics in memory of Pierre Delattre (pp. 31-44). The Hague: Mouton.

Abramson, A. S., & Lisker, L. (1965). Voice onset time in stop consonants: Acoustic analysis and synthesis. In D. E. Commins (Ed.), Proceedings of the 5th International Congress of Acoustics (A-51). Liege: G. Thone.

Bailey, P. J., Summerfield, A. Q., & Dorman, M. (1977). On the identification of sine-wave analogues of certain speech sounds. Haskins Laboratories Status Report on Speech Research, SR-51/52, 1-25.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. Perception & Psychophysics, 29, 191-211.

Catford, J. C. (1977). Fundamental problems in phonetics. Bloomington: Indiana University Press.

Collier, R., & 't Hart, J. (1975). The role of intonation in speech perception. In A. Cohen & S. E. G. Nooteboom (Eds.), Structure and process in speech perception (pp. 107-123). New York: Springer-Verlag.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of America, 27, 769-773.

Fant, C. G. M. (1956). On the predictability of formant levels and spectrum envelopes from formant frequencies. In M. Halle, H. Lunt, & H. MacLean (Eds.), For Roman Jakobson (pp. 109-120). The Hague: Mouton.

Fry, D. B. (1958). Experiments in the perception of stress. Language and Speech, 1, 126-152.

Greenberg, S. (1980). Temporal neural coding of pitch and vowel quality. UCLA Working Papers in Phonetics, 52, 1-183.

Grunke, M. E., & Pisoni, D. B. (1982). Perceptual learning of mirror-image acoustic patterns. Perception & Psychophysics, 31, 210-218.

Hadding-Koch, K., & Studdert-Kennedy, M. (1964). An experimental study of some intonation contours. Phonetica, 11, 175-185.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America, 25, 105-113.

Joos, M. (1948). Acoustic phonetics. Language, 24 (Suppl., Language Monographs 23), 1-136.

Lea, W. A. (1979). Testing linguistic stress rules with listeners' perception. In J. J. Wolf & D. H. Klatt (Eds.), Speech communication papers (pp. 415-418). New York: Acoustical Society of America.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. Glossa, 7, 107-122.

Lehiste, I., & Peterson, G. E. (1959). Vowel amplitude and phonemic stress in American English. Journal of the Acoustical Society of America, 31, 428-435.

Lehiste, I., & Peterson, G. E. (1961). Some basic considerations in the analysis of intonation. Journal of the Acoustical Society of America, 33, 419-423.

Licklider, J. C. R. (1956). Auditory frequency analysis. In C. Cherry (Ed.), Information theory. New York: Academic.

Lieberman, P. (1965). On the acoustic basis of the perception of intonation by linguists. Word, 20, 40-54.

Lieberman, P. (1967). Intonation, perception and language (Research Monograph No. 38). Cambridge: MIT Press.

Meyer-Eppler, W. (1957). Realization of prosodic features in whispered speech. Journal of the Acoustical Society of America, 29, 104-106.

Morton, J., & Jassem, W. (1965). Acoustic correlates of stress. Language and Speech, 8, 159-181.

Plomp, R. (1967). Pitch of complex tones. Journal of the Acoustical Society of America, 41, 1526-1533.

Remez, R. E., Rubin, P. E., & Carrell, T. D. (1981). Phonetic perception of sinusoidal signals: Effects of amplitude variation. Journal of the Acoustical Society of America, 69, S114.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. Science, 212, 947-950.

Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. Journal of the Acoustical Society of America, 42, 191-198.

Schouten, J. F. (1940). The residue and the mechanism of hearing. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, 43, 991-999.

Schwab, E. C. (1981). Auditory and phonetic processing for tone analogs of speech. Unpublished doctoral dissertation, State University of New York at Buffalo.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. Journal of the Acoustical Society of America, 64, 1582-1592.

Summerfield, Q., & Haggard, M. (1977). On the dissociation of spectral and temporal cues to the voicing distinction in initial stop consonants. Journal of the Acoustical Society of America, 62, 435-448.

Footnotes

[1]A pure tone is not a formant. A sinusoid is defined by the function $y = a\sin x$, and may occur at any frequency within the audible range. A formant is a natural resonance of the vocal tract, and its frequency is defined as the peak of the spectrum envelope drawn to enclose the harmonics produced by the excitation of the vocal tract (Fant, 1956). Although we have constructed sinusoids that imitate the pattern of formant center-frequency variation, they do not also imitate the acoustic structure of formants, by this definition. For a basic discussion of the physical acoustics of speech, see Joos (1948).

[2]The intonation of a sentence is its pitch contour (Catford, 1977), though this definition is perceptually troublesome. This is so because the term pitch is traditionally used to refer to that perceptual impression correlated with fundamental frequency. Intonation is also correlated mainly with fundamental frequency, although pitch applies to speech and nonspeech, and intonation more narrowly applies to speech exclusively. In view of this,

is sentence intonation the product or the equivalent of sentence pitch contour? The fact that aspects of signal duration and power intrude on the perception of both intonation and pitch argues that both terms name the same attribute. The influence of lexical structure in judging sentence melody argues against any simple equivalence, although it by no means warrants that a separate auditory impression of pitch contributes to the impression of intonation. (Linguists have occasionally combined the analysis of intonation and word stress [reviewed by Lieberman, 1967], although to do so does not dismiss the phenomenon of sentence pitch--it simply adds another problem to consider.) Our present use of the term, then, refers to the fact that sentence "pitch contour," sentence "melody," and sentence "intonation" seem to indicate the same aspect of spoken sentences, although its perceptual derivation is difficult to resolve.

PUBLICATIONS
APPENDIX

PUBLICATIONS

Abramson, A. S., & Lisker, L. (in press). Relative power of cues: F0 shift vs. voice timing. In V. Fromkin (Ed.), Phonetic linguistics. New York: Academic Press.

Alfonso, P. J., Watson, B. C., & Baer, T. (1984). Muscle, movement, and acoustic measurements of stutterers' laryngeal reaction times. In M. Edwards (Ed.), Proceedings of the 19th Congress of the International Association of Logopedics and Phoniatrics (Vol. II, pp. 580-585). Perth, Scotland: Dansoot Print Limited.

Baer, T., & Alfonso, P. J. (1984). On simultaneous neuromuscular, movement, and acoustic measures of speech articulation. In R. G. Daniloff (Ed.), Articulation assessment and treatment issues (pp. 195-214). San Diego: College-Hill Press.

Bentin, S., Bargai, N., & Katz, L. (1984). Orthographies and phonemic coding for lexical access: Evidence from Hebrew. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10.

Borden, G. (1984). Consideration of motor-sensory targets and a problem in perception. In H. Winitz (Ed.), Treating articulation disorders: For clinicians by clinicians (pp. 51-65). Baltimore: University Park Press.

Browman, C. P., & Goldstein, L. M. (in press). Dynamic modeling of phonetic structure. In V. Fromkin (Ed.), Phonetic linguistics. New York: Academic Press.

Crowder, R. G. (1984). Is it just reading? Comments on the papers by Mann, Morrison, and Wolford and Fowler. Developmental Review, 4, 48-61.

Crowder, R. G., & Repp, B. H. (1984). Single formant contrast in vowel identification. Perception & Psychophysics, 35, 372-378.

Hanson, V. L., Liberman, I. Y., & Shankweiler, D. (1984). Linguistic coding by deaf children in relation to beginning reading success. Journal of Experimental Child Psychology, 37, 378-393.

Harris, K. S. (1984). Coarticulation as a component in articulatory description. In R. G. Daniloff (Ed.), Articulation assessment and treatment issues (pp. 147-167). San Diego, CA: College Hill Press.

Harris, K. S., Tuller, B., & Kelso, J. A. S. (in press). Temporal invariance in the production of speech. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: Erlbaum.

Hawkins, S. (1984). On the development of motor control in speech: Evidence from studies of temporal coordination. In N. J. Lass (Ed.), Speech and language: Research and theory (Vol. 11, pp. 317-373). New York: Academic Press.

Hoffman, P. R., Daniloff, R. G., Alfonso, P. J., & Schuckers, G. H. (1984). Multiple phoneme misarticulating children's perception and production of voice onset time. Perceptual and Motor Skills, 58, 603-610.

Kelso, J. A. S. (in press). Remarks on Preparatory processes. In E. Donchin (Ed.), Cognitive psychophysiology: Carmel 1. Hillsdale, NJ: Erlbaum.

Kelso, J. A. S. (in press). Phase transitions and critical behavior in human bimanual coordination. American Journal of Physiology: Regulatory, Integrative and Comparative.

Kelso, J. A. S., & Tuller, B. (1984). Converging evidence in support of common dynamical principles for speech and movement coordination. American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 246, R928-R935.

Kelso, J. A. S., Tuller, B., & Harris K. S. (in press). Authors' response: A theoretical note on speech timing. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: Erlbaum.

Kugler, P. N., Turvey, M. T., Carello, C., & Shaw, R. (in press). The physics of controlled collisions: A reverie about locomotion. In W. H. War-

ren, Jr. & R. E. Shaw (Eds.), Persistence and change: Proceedings of the First International Conference on Event Perception. Hillsdale, NJ: Erlbaum.

Liberman, A. M. (in press). Brief comments on invariance in phonetic perception. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: Erlbaum.

Lisker, L. (in press). Review (The phonetics of Macedonian, by N. Minissi, N. Kitanovski, & U. Cinque). Journal of the Acoustical Society of America.

Lisker, L., & Baer, T. (in press). Laryngeal management at utterance-internal word boundary in American English. Language and Speech.

Mann, V. A. (in press). Prediction and prevention of early reading difficulty. Annals of Dyslexia.

Mattingly, I. G. (1984). Reading, linguistic awareness and language acquisition. In J. Downing & R. Valtin (Eds.), Language awareness and learning to read (pp. 9-25). New York: Springer-Verlag.

Morrongiello, B. A., Robson, R. C., Best, C. T., & Clifton, R. K. (1984). Trading relations in the perception of speech by 5-year-old children. Journal of Experimental Psychology, 37, 231-250.

Rakerd, B. (1984). Vowels in consonantal context are perceived more linguistically than are isolated vowels: Evidence from an individual differences scaling study. Perception & Psychophysics, 35, 123-136.

Rakerd, B., Verbrugge, R. R., & Shankweiler, D. (in press). Monitoring for vowels in isolation and in a consonantal context. Journal of the Acoustical Society of America.

Recasens, D. (1984). V-to-C coarticulation in Catalan VCV sequences: An articulatory and acoustical study. Journal of Phonetics, 12, 61-73.

Recasens, D. (in press). Vowel-to-vowel coarticulation in Catalan VCV sequences. Journal of the Acoustical Society of America.

Remez, R. E., & Rubin, P. E. (1984). On the perception of intonation from sinusoidal sentences. Perception & Psychophysics, 35, 429-440.

Repp, B. H. (1984). The role of release bursts in the perception of [s]-stop clusters. Journal of the Acoustical Society of America, 75, 1219-1230.

Repp, B. H. (in press). Effects of temporal stimulus properties on perception of the [sl]-[spl] distinction. Phonetica.

Repp, B. H. (in press). Closure duration and release burst amplitude cues to stop consonant manner and place of articulation. Language and Speech.

Repp, B. H., & Liberman, A. M. (in press). Phonetic category boundaries are flexible. In S. N. Harnad (Ed.), Categorical perception. New York: Cambridge University Press.

Serafine, M. L., Crowder, R. G., & Repp, B. H. (in press). Integration of melody and text in memory for songs. Cognition.

Studdert-Kennedy, M. (in press). Sources of variability in early speech development. In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: Erlbaum.

Studdert-Kennedy, M. (in press). Invariance: Functional or descriptive? In J. S. Perkell & D. H. Klatt (Eds.), Invariance and variability of speech processes. Hillsdale, NJ: Erlbaum.

Tuller, B. (in press). On categorizing aphasic speech errors. Neuropsychologia.

Tuller, B., & Kelso, J. A. S. (in press). The timing of articulatory gestures: Evidence for relational invariance. Journal of the Acoustical Society of America.

Turvey, M. T., Feldman, L. B., & Lukatela, G. (1984). The Serbo-Croatian orthography constrains the reader to a phonologically analytic strategy. In L. Henderson (Ed.), Orthographies and reading. London: Erlbaum.

APPENDIX

Status Report		DTIC	ERIC
SR-21/22	January - June 1970	AD 719382	ED 044-679
SR-23	July - September 1970	AD 723586	ED 052-654
SR-24	October - December 1970	AD 727616	ED 052-653
SR-25/26	January - June 1971	AD 730013	ED 056-560
SR-27	July - September 1971	AD 749339	ED 071-533
SR-28	October - December 1971	AD 742140	ED 061-837
SR-29/30	January - June 1972	AD 750001	ED 071-484
SR-31/32	July - December 1972	AD 757954	ED 077-285
SR-33	January - March 1973	AD 762373	ED 081-263
SR-34	April - June 1973	AD 766178	ED 081-295
SR-35/36	July - December 1973	AD 774799	ED 094-444
SR-37/38	January - June 1974	AD 783548	ED 094-445
SR-39/40	July - December 1974	AD A007342	ED 102-633
SR-41	January - March 1975	AD A013325	ED 109-722
SR-42/43	April - September 1975	AD A018369	ED 117-770
SR-44	October - December 1975	AD A023059	ED 119-273
SR-45/46	January - June 1976	AD A026196	ED 123-678
SR-47	July - September 1976	AD A031789	ED 128-870
SR-48	October - December 1976	AD A036735	ED 135-028
SR-49	January - March 1977	AD A041460	ED 141-864
SR-50	April - June 1977	AD A044820	ED 144-138
SR-51/52	July - December 1977	AD A049215	ED 147-892
SR-53	January - March 1978	AD A055853	ED 155-760
SR-54	April - June 1978	AD A067070	ED 161-096
SR-55/56	July - December 1978	AD A065575	ED 166-757
SR-57	January - March 1979	AD A083179	ED 170-823
SR-58	April - June 1979	AD A077663	ED 178-967
SR-59/60	July - December 1979	AD A082034	ED 181-525
SR-61	January - March 1980	AD A085320	ED 185-636
SR-62	April - June 1980	AD A095062	ED 196-099
SR-63/64	July - December 1980	AD A095860	ED 197-416
SR-65	January - March 1981	AD A099958	ED 201-022
SR-66	April - June 1981	AD A105090	ED 206-038
SR-67/68	July - December 1981	AD A111385	ED 212-010
SR-69	January - March 1982	AD A120819	ED 214-226
SR-70	April - June 1982	AD A119426	ED 219-834
SR-71/72	July - December 1982	AD A124596	ED 225-212
SR-73	January - March 1983	AD A129713	ED 229-816
SR-74/75	April - September 1983	AD A136416	ED 236-753
SR-76	October - December 1983	AD A140176	ED 241-973
SR-77/78	January - June 1984	**	**

Information on ordering any of these issues may be found on the following page.

DTIC and/or ERIC order numbers not yet assigned.

AD numbers may be ordered from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22151

ED numbers may be ordered from:

ERIC Document Reproduction Service
Computer Microfilm International
Corp. (CMIC)
P.O. Box 190
Arlington, Virginia 22210

Haskins Laboratories Status Report on Speech Research is abstracted in Language and Language Behavior Abstracts, P.O. Box 22206, San Diego, California 92122.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Haskins Laboratories
270 Crown Street
New Haven, CT 06511

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
N/A

3. REPORT TITLE
Haskins Laboratories Status Report on Speech Research, SR-77/78 (1984)

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Interim Scientific Report

5. AUTHOR(S) (First name, middle initial, last name)
Staff of Haskins Laboratories, Alvin M. Liberman, P.I.

6. REPORT DATE
August, 1984

7a. TOTAL NO. OF PAGES
230

7b. NO. OF REFS
429

8a. CONTRACT OR GRANT NO.

HD-01994 NS13870
HD-16591 NS13617
N01-HD-1-2420 NS18010
RR-05596 N00014-83-K-0083
BNS-8111470

9a. ORIGINATOR'S REPORT NUMBER(S)
SR-77/78

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
None

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited*

11. SUPPLEMENTARY NOTES
N/A

12. SPONSORING MILITARY ACTIVITY
See No. 8

13. ABSTRACT

This report (1 January-30 June) is one of a regular series on the status and progress of studies on the nature of speech, instrumentation for its investigation, and practical applications. Manuscripts cover the following topics:

Sources of variability in early speech development
Invariance: Functional or descriptive?
Brief comments on invariance in phonetic perception
Phonetic category boundaries are flexible
On categorizing aphasic speech errors
Universal and language particular aspects of vowel-to-vowel coarticulation
Functionally specific articulatory cooperation following jaw perturbation during speech: Evidence for coordinative structures
Formant integration and the perception of nasal vowel height
Relative power of cues: F0 shift vs. voice timing
Laryngeal management at utterance-internal word boundary in American English
Closure duration and release burst amplitude cues to stop consonant manner and place of articulation
Effects of temporal stimulus properties on perception of the [sl]-[spl] distinction
The physics of controlled conditions: A reverie about locomotion
On the perception of intonation from sinusoidal sentences

→ over

DD FORM 1 NOV 65 1473 (PAGE 1)

S/N 0101-807-6811

UNCLASSIFIED

Security Classification

A-31408

*This document contains no information not freely available to the general public. It is distributed primarily for library use.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Speech Perception; phonetic, invariance cues, F0 shift, timing nasal vowels, formant integration phonetic, category boundaries stop consonant, closure, release intonation, sinewave synthesis phonetic, temporal stimulus properties						
Speech Articulation; aphasia, error categories larynx, word boundaries coarticulation, vowel-to-vowel						
Motor Control; articulation, jaw perturbation, coordinative structures locomotion, controlled conditions						
Speech Development; children, variability						

END

FILMED

10-84

DTIC